UNIVERSITY
OF TASMANIA

Genetic association analysis of *indica* rice yield and related traits in irrigated ecosystems

by

Shanshan Liang

M.Sc. (Chinese Academy of Agricultural Science, China)

B.Sc. (Shandong Agricultural University, China)

Submitted in fulfilment of the requirements for the Doctor of Philosophy

University of Tasmania

June 2016

Declarations

The thesis contains no material which has been accepted for the a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

This thesis may be made available for loan and limited copying in accordance with the *Copyright Act* 1968.

Student Signature:

Primary Supervisor Signature:

Statement of co-authorship

This thesis was completed during the course of my enrolment in a PhD degree in the School of Land and Food at the University of Tasmania. This thesis contains no experimental results that have previously presented for any degree at this or other institution.

This thesis contains one literature review chapter and three main research chapters. Results described in the three research chapters have been published (Chapter 3 and Chapter 4) and accepted (Chapter 5) for publication.

The following people and institutions contributed to the publication of work undertaken as part of this thesis:

Candidate	Shanshan Liang, University of Tasmania
Author 1	Guoyou Ye, International Rice Research Institute
Author 2	Meixue Zhou, University of Tasmania
Author 3	David McNeil, University of Tasmania
Author 4	Xiangqian Zhao, International Rice Research Institute
Author 5	Guangjun Ren, Sichuan Academy of Agricultural Sciences, China
Author 6	Jianping Liu, Pingxiang Institute of Agricultural Sciences, China
Author 7	Chenxiao Sun, China National Rice Research Institute, China
Author 8	Liwen Wu, China National Rice Research Institute, China

Paper 1: Liang, S., Ren, G. Liu, J., Zhao, X., Zhou, M., McNeil, D. and Ye, G. 2015 Genotype-by-environment interaction is important for grain yield in irrigated lowland rice. Field Crops Research. 180:90-99. **---Located in Chapter 3**

Candidate conducted 70% of the experimental work and phenotype data recording and drafted the paper.

Author 1 contributed to the conception and design of the project. Author 2 and Author 3 contributed to the proof-reading and revision of the manuscript. Author 4, Author 5 and Author 6 and contributed to the experimental work and phenotype data collection.

Paper 2: Liang, S., Sun, C., Ren, G., Zhao, X., Zhou, M., McNeil, D. and Ye, G. 2016. Usefulness of the cloned and fine-mapped genes/QTLs for grain yield and related traits in *indica* rice breeding for irrigated ecosystems. Field Crops Research. 187:58-73. ---**Located in Chapter 4.**

Candidate conducted 70% of the experimental work and phenotype data recording and drafted the paper.

Author 1 contributed to the conception and design of the project and interpretation of part of the research data. Author 2 and Author 3 contributed to the proof-reading and revision of the manuscript. Author 4, Author 5 and Author 7 contributed to the experimental work and phenotype data collection.

Paper 3: Liang, S., Wu, L., Ren, G., Zhao, X., Zhou, M., McNeil, D. and Ye, G. 2016 Genome-wide association study of grain yield and related traits using a collection of advanced *indica* rice breeding lines for irrigated ecosystems. Field Crops Research. In Press. ---**Located in Chapter 5**

Candidate contributed to 70% of this work, including experimental work, data analysis and paper drafting.

Author 1 contributed to the conception and design of the project and interpretation of part of the research data. Author 2 and Author 3 contributed to the proof-reading and revision of the manuscript. Author 4, Author 5 and Author 8 contributed to the experimental work and phenotype data collection.

We agree with the above stated "contribution of work undertaken" for each of the above published (or submitted) peer-reviewed manuscripts contributing to this thesis:

Student Signature:

Primary Supervisor Signature:

Acknowledgements

During the course of this project I was financially supported by University of Tasmania, International Rice Research Institute (IRRI), and Monsanto's Beachell-Borlaug International Scholars Program (MBBISP). This research would not have been possible without the supervision and guidance from my excellent supervisory team. I would like to give special thanks to all of my supervisors for their commitment and dedication to providing the full support, inspiration, discussion and encouragement. I am very thankful to Dr. Guoyou Ye, who helped to find fund to support my PhD research. He was the major drive to get the MBBISP. I also thank my application referees, Drs. Achim Dobermann, former Deputy Director General for Research of IRRI and Xianchun Xia of the Chinese Academy of Agricultural Science, for their full support. Dr. Ye designed experiments, coordinated to get all the plant materials and deployed the field experiments and set up our molecular lab from scratch. He always inspired me with his earnest and persistence for scientific research. His attitude to scientific research, integrity, dedication and diligence impressed me a lot. It's a great privilege to work with a great scientist like Dr. Ye. Many sincere thanks to A/Prof Meixue Zhou and Prof. David McNeil for their patience and generousness to help me with writing manuscript and thesis. Dr. Zhou is knowledgeable and hard working. I learnt a lot from him and really appreciate him. Dr. McNeil helped me to get the fee waiver from UTAS. He also provided valuable suggestions and comments on my manuscripts and thesis. I truly thank him for his knowledge and patience. All my supervisors helped me with great patience, knowledge and experience. I couldn't thank them more.

My special thanks go to Dr. Guangjun Ren at the Sichuan Academy of Agricultural Sciences (Chengdu, China) and Mr. Jianping Liu at the Pingxiang Institute of Agricultural Sciences (Jiangxi, China), who contribute to the field testing experiment in China. I also thank Ms. Violeta Bartolome for helping with data analysis and Dr. Ed Redoña for allowing me to use their molecular lab facilities.

My lab mates from Dr. Ye's team, e.g. Drs. Xiangqian Zhao, Zhaoming Chen, Fang Wang, Mao Qin, Lijun Meng, Xiang Lu, Xiufang Ma, Min Li and Yunlong Pang helped with the field trials, lab techniques or data analysis. I am so grateful for their generous support and casual discussions. Technicians Kimberly Ponce and Severino Marundon helped me with field trials and lab experiment. I thank them for their full support. I also would like to thank all my friends from IRRI and UTAS, who made my life entertaining and enjoyable. I won't

list all your names but I will always remember you and the time we spent together. Thank you all for being such good company in this journey.

I would like to thank sincerely all the people from IRRI and School of Land and Food of UTAS for their administrative support. I also thank Dr. Darshan Brar from IRRI, who was my co-supervisor at IRRI and retired before I finish my degree. I wish him an enjoyable retiring life.

My thank goes to Prof. Qingqi Li from Shandong Agricultural University, who has been supporting me financially and spiritually since I was an undergraduate. His encouragement and support inspired me to pursuit PhD degree. I also would like to thank my Master supervisor Dr. Zhonghu He from the Chinese Academy of Agricultural Science for his continuing support and encouragement.

Last but not least, special thanks to my dearest husband and daughter, Gaohua Niu and Shuwei Niu, for their consideration and support while I am away from home such a long time. I love you two more than I can say. I feel sorry for my daughter that I should spend more time with her. She was doing well when I left her and backed to school. I am so proud of my dearest little daughter. I thank my dearest husband for dedicating to take care of our daughter without complaining. Without his sacrifice, I couldn't focus on my study and complete my degree. I also got support from my father-in-law, my parents, my brother and sister-in-law. I am really thankful for their understanding and supporting.

Abstract

Rice is the dominant food source for more than half of the world's population. Irrigated rice predominates in global rice production as well as in most rice producing countries. Further improvement of grain yield (GY) by combining well-proven conventional breeding methods with new techniques offered by modern molecular biology and genomics is urgently needed to feed the growing world population. The studies reported in this thesis are part of the effort led by IRRI to break the yield barrier for irrigated rice. This research provides essential phenotypic and genetic information directly relevant to future breeding for irrigated ecosystems. The specific objectives are: i) to gain a better understanding of the importance of genotype-environment interaction (GEI) on GY in irrigated rice ecosystems; ii) to test the usefulness of 39 fine-mapped or cloned genes/QTLs for GY and yield related traits in a breeding population through association analysis and iii) to identify new markers associated with GY and related traits through a genome-wide association study (GWAS).

The studies were conducted using a collection of 392 genotypes including released cultivars and advanced lines from several large breeding programs worldwide. These are being used as parental lines in IRRI's recurrent selection and variety development programs. Field trials were conducted in eight environments including Jiangxi (JX) and Sichuan (SC) in China, and six season (2) and nitrogen rate (3) combinations at IRRI headquarters (Philippines). The two seasons were the dry season (DS) and wet season (WS) of 2012. For the DS the three nitrogen rates were no nitrogen, 90 kg ha^{-1} and 180 kg ha^{-1}, designed as DS1, DS2 and DS3. For the WS the three nitrogen rates were no nitrogen, 45 kg ha^{-1} and 90 kg ha^{-1}, designed as WS1, WS2 and WS3. GY and the following 10 traits were measured, including grain number per panicle (GN), panicle number per plant (PN), thousand grain weight (TGW), days to flowering (DTF), primary branch per panicle (PB), plant height (PH), secondary branch per panicle (SB), Spikelet number per panicle (SN), seed setting rate (SR) and tiller number (TN). All the 11 traits were tested in DS1, DS3, WS1, WS2 and WS3, while selected traits were tested in DS2, JX and SC.

A wide range of variations across genotypes and environments were observed for all traits. Genotype, environment and GEI all significantly affected GY and yield related traits. GEI was more important than the genotypic main effect for GY, SR and PN but less important for other traits. For GY, the genotype-by-season interaction and genotype-by-season-by-nitrogen interaction were more important than the genotype-by-nitrogen interaction. The genotypes were clustered into 10 groups using an agglomerative hierarchical clustering procedure. The eight environments were grouped into three groups using the biplot

of the additive main effects and multiplicative interaction (AMMI) analysis. The three environments (nitrogen rates) in the WS and SC were grouped together (E1) and three environments (nitrogen rate) in the DS formed another group (E2). JX alone was the third group (E3). JX was relatively closer to the IRRI DS environments in the biplot. It indicated that IRRI breeding lines with stable and good performance in the WS could be used in SC (directly as varieties or as parental lines in breeding) and that selection is better done in the DS in IRRI for use in JX, China. Great attention should be paid to the relevance of performance at IRRI to their target production environments when IRRI breeding lines are introduced. Breeding for DS and WS separately at IRRI was recommended to exploit the repeatable GEI caused by seasonal variation.

To test the usefulness of fine-mapped or cloned genes/QTLs for GY and yield related traits, 46 molecular markers tightly linked to the chosen 39 genes/QTLs were used to genotype 360 of the 392 lines. Population structure analysed with 53 SSR markers evenly distributed on all chromosomes using STRUCTURE program indicated that the whole population could be divided into two subpopulations of 205 and 155 lines. A mixed linear model incorporating genetic relatedness between genotypes was chosen by comparing four commonly used statistical models. The selected model was used to conduct association analysis for all the tested traits in each of the eight testing environments and the average environment defined as the average across the testing environments. Analyses were separately carried out for the whole population and the two subpopulations. All the 39 target genes/QTLs were associated with two or more measured traits including traits not previously reported. *GW6* and *Gn1a* were associated with nine and eight traits, respectively. *Ghd7*, *qSPP7*, *SCM2* and *SPP1* were associated with seven traits, *GIF1* and *Ltn* were with six traits, *GS3*, *GW2*, *gw3.1*, *htd1*, *Nop(t)*, *qGY2-1* and *qPH6-1* with five traits, *D10*, *d27*, *DEP2*, *DWL1*, *Gnp4*, *Gw1-1*, *GW3*, *gw5*, *MOC1*, *PAP2*, *qGL7*, *qGL7-2* and *qGN4-1* with four traits, *D88*, *Ghd8*, *GS5*, *Gw1-2*, *IPA1*, *qSH3* and *RPH* with three traits and *ep3*, *gw8.1*, *gw9.1* and *qPDS3* with two traits. A total of 16 genes/QTLs were found to be associated with GY. *GS3*, *GW1-1* and *d27* were associated with GY in two testing environments and the others were only in one. For the three yield component traits GN, PN and TGW, there were 16, six and 10 genes/QTLs identified to be associated with in one or more environments. Eleven genes/QTLs were associated with SN and SR, respectively. There were 29 and six genes/QTLs detected to be associated with PB and SB, respectively. TN had the least number of significant genes/QTLs associated, which were five. All the 39 genes/QTLs were associated with PH in one testing environment or the average environment. There were 25 genes/QTLs found to be associated with DTF in one or more environments. Significant gene-

by-environment interaction was present for all the studied genes/QTLs. However, GY could not be well predicted using the markers significantly associated with measured traits or all the target markers based on stepwise multiple linear regression analysis. The adjusted coefficient of determination (R^2) ranged from 0.024 to 0.191 for the final selected models considering the associated markers only and from 0.039 to 0.261 for the final selected models considering all the target markers. Nevertheless the known genes might be explicitly utilized in developing more efficient selection criteria for enhancing selection accuracy.

To identify new markers associated with GY and related traits, 327 of the 392 lines were genotyped with SNPs using the genotyping-by-sequencing method. Model based population structure analysis was conducted with a subset of 1072 evenly distributed SNPs. The results indicated that the likely number of subpopulations was two, with subpopulation 1 consisting of 234 lines and subpopulation 2 consisting of 93 lines. There were 56 common lines between the two smaller subpopulations derived from the population structure analysis results using SSR and SNP. Based on comparison of multiple models for selected trait and environment combinations, the PK model implemented in TASSEL software using principle components to correct population structure and the relative kinship matrix to adjust the unequal familial relatedness between the individuals was selected to conduct the GWAS for all the traits in all the testing environments. A total of 452 marker-trait associations that were delineated into 43 QTLs were identified for all traits but PB, SB and SR with 39 QTLs being not reported before. Three QTLs were identified for GY on chromosome 6, 9 and 12 but only in DS2. The numbers of QTLs identified for PN, GN and TGW were 26, four and two. There were nine and four QTLs detected for DTF and TN, respectively. Two QTLs were identified for PH and SN, respectively. Most of the detected QTLs were found in only one environment. One of the QTL for DTF on chromosome 3 was identified in multiple environments and corresponds to *Hd9* reported in previous studies. Two of the QTLs for PN on chromosome 1 were in the regions of previously fine-mapped QTLs, *Gw1-1* and *Gw1-2*, for TGW. The effects of newly identified QTLs were relatively small with the highest percentage of variance explained by a single QTL being 9.6%.

Gene-by-environment interaction, pleiotropy and small effects of the well characterized genes/QTLs and newly detected QTLs imply that selection accuracy using the identified genes/QTLs is low. Improving yield and related quantitative traits through marker-assisted selection remains a big challenge. Recently developed genomic selection that utilizes markers in linkage disequilibrium with all genes affecting trait of interest and captures interactions between genes should be exploited.

Publications arising from this thesis

1. **Liang, S.**, Ren, G. Liu, J., Zhao, X., Zhou, M., McNeil, D. and Ye, G. 2015. Genotype-by-environment interaction is important for grain yield in irrigated lowland rice. Field Crops Research. 180:90-99. (Chapter 3)
2. **Liang, S.**, Sun, C., Ren, G., Zhao, X., Zhou, M., McNeil, D. and Ye, G. 2016. Usefulness of the cloned and fine-mapped genes/QTLs for grain yield and related traits in *indica* rice breeding for irrigated ecosystems. Field Crops Research. 187:58-73. (Chapter 4)
3. **Liang, S.**, Wu, L., Ren, G., Zhao, X., Zhou, Z., McNeil, D. and Ye, G. 2016 Genome-wide association study of grain yield and related traits using a collection of advanced *indica* rice breeding lines for irrigated ecosystems. Field Crops Research. In Press. (Chapter 5)

Table of Contents

List of tables

List of figures

List of abbreviations

Abbreviation	Full name
AAC	apparent amylose content
AC	amylose content
Al	Aluminium
AM	association mapping
BLUE	best linear unbiased estimator
BSA	bulk segregant analysis
BSC	bundle sheath cells
CIM	composite interval mapping
cM	cent Morgan (Genetic distances)
CRISPR	clustered regularly interspaced short palindromic repeats
DH	doubled haploid
DS	dry season
DSB	double-strand breaks
DTF	days to flowering
FT	Flowering time
GBLUP	genomic best linear unbiased prediction
GBS	genotyping-by-sequencing
GEBV	genomic estimated breeding value
GEI	genotype-by-environment interaction
GLM	general linear model
GM	genetically modification
GN	grain number per panicle
GPD	gametic phase disequilibrium
GS	genomic selection
GTA	gene/QTL trait associations
GW	grain width
GWAS	genome-wide association study
GY	grain yield
HD	heading date
HDR	homology directed repair
IRRI	International Rice Research Institute
JX	Jiangxi, China
LD	linkage disequilibrium
LE	linkage equilibrium
MABC	marker-assisted backcrossing
MAF	minor allele frequency
MARS	marker-assisted recurrent selection
MAS	marker-assisted selection
MC	mesophyll cells
MET	multi-environment trials
MLM	mixed linear model

MLR	multiple linear regression
mM	Milimolar
MSD	mean of the squared differences
MTA	marker-trait association
NAM	designed mapping populations
ng	nano gram
NGS	next-generation sequencing
NHEJ	nonhomologous end joining
NJ	neighbour joining
PB	number of primary branches per panicle
PCA	principal component analysis
PH	plant height
PIC	polymorphism information content
PN	panicle number per plant
PT	pasting temperature
QEI	QTL-by-environment interaction
QQ plot	quantile-quantile plot
RIL	recombinant inbred line
RRBLUP	ridge regression best linear unbiased prediction
RS	recurrent selection
SB	number secondary branches per panicle
SC	Sichuan, China
ShB	sheath blight
SN	spikelet number per panicle
SR	seed setting rate
SSN	sequence-specific nucleases
STS	sequence tagged site
TALEN	transcription activator-like effector nucleases
TGW	thousand grain weight
TILLING	targeting-induced local lesions in genomes
TN	tiller number per plant
TPE	target population of environments
WS	wet season

Chapter 1 General introduction

1.1 Importance of rice (*Oryza sativa* L.)

Rice (*Oryza sativa* L.) is one of the most important crops in the world having played a central role in human nutrition and cultured for nearly10,000 years (Molina et al., 2011). Rice is grown under a wide range of agronomic conditions providing a staple food for more than half of the world's population (Delseny et al., 2001). It is estimated that the global rice production must reach 800 million tonnes in 2025 (currently world paddy production in 2014 is 744.4 million tons) (FAOSTAT, 2014) to meet the demand for rice consumption. This extra rice will have to be produced using less land, less water, less labour and fewer chemical inputs. Therefore increasing rice production is still a big challenge today.

Rice also serves as a monocot model system in plant genomics benefiting from having the smallest genome of major cereals, dense genetic maps, relative ease of genetic transformation and abundant genetic and genomic resources, including mutants, cultivated landraces and wild species (Paterson et al., 2005; Shimamoto and Kyozuka, 2002; Xu et al., 2005). The extensive genome colinearity among the Poaceae made rice the model organism for the cereal grasses. The genomic sequences of rice were completely determined by the International Rice Genome Sequencing project in 2004 (IRGSP, 2005). Rice was the second species, after *Arabidopsis*, among the seed plants to have its genome well understood. The *Agrobacterium*-mediated rice transformation method allows analyzing the functions of the gene of interest by producing transgenic rice. Progress in molecular techniques and tools make gene isolation and function analysis possible. Map-based cloning methods as well as tagging methods using endogenous transposons and exogenous elements have been used to isolate many important genes (Peters et al., 2003; Remington et al., 2001). Reverse genetic studies, such as screening of knockout lines, making knockout lines and using the targeting-induced local lesions in genomes (TILLING) method are possible. Genome wide expression analysis such as microarray and massively parallel signature sequencing (MPSS) for rice research are available and will provide a wealth of functional genomic information in rice.

1.2 Breeding targets of rice

The Green Revolution in the 1960s greatly increased rice production widely around the world. However, the production potential of modern rice cultivars has remained stagnant for the past several decades (Nguyen and Ferrero, 2006). Biotic and abiotic stresses, as well as narrow

genetic diversity among present breeding lines are the major constraints to further increases in rice productivity (Ali et al., 2006; Niño-Liu et al., 2006; Thomson et al., 2010) Domestication and modern breeding has narrowed down the rice genetic diversity significantly. Modern cultivated rice is estimated to retain only approximately 10-20% of the genetic diversity present in its wild rice ancestor, *O. rufipogon* (Caicedo et al., 2007; Zhu et al., 2007).

At least 50% of the increase of rice production in the past has been due to the adoption of new cultivars. Rice production in China has increased more than three times over the past six decades due to increased grain yield (GY) per hectare rather than increased planting area (Yu et al., 2012). The increases in rice yield in China between 1960 and 1980 owed a lot to a combination of increased N application to N responsive hybrid varieties. However, the 52.6% rice yield increase in China since 1980 is comprised of increases that can be attributed to climate (4.4%), management (9.3%) and predominantly variety (38.9%). Genetic improvement is the decisive factor and contributed 74.0% of the total increase in yield using the Agro-C model and census yields to investigate how climate, crop management and variety renewal have interactively affected the rice yields in China for the past three decades (Yu et al., 2012). It is expected that breeding will still be one of the key practices for further increasing productivity by improving yield potential and stability. Conventional breeding practice for genetic improvement of a self-pollinated crop such as rice generally start with crossing between two cultivars which usually are homozygous for some desirable traits. The conventional breeding methods, such as pedigree, bulk and backcross, have some disadvantages (Fujimaki, 1980), e.g. limited use of the full range of available genetic resources, restrictions of the potential for genetic recombination and difficulty of continuing to obtain improvements in successive breeding cycles. Therefore, the conventional procedures in plant breeding have caused a severe reduction in the genetic diversity of the modern crop cultivars (Cuevas-Pérez et al., 1992; Mishra, 2002; Montalban et al., 1998; Rai, 2002).

1.3 Genotype-by-environment interaction

The performance of a rice cultivar is the collective effect of its genotype, the environment where it is grown and the interaction between the genotype and environment (Gomez and Gomez, 1984). The term genotype-by-environment interaction (GEI) is used to indicate fluctuation in the performance of genotypes or a change in the magnitude of the difference

between genotypes under multiple environments. The exhibited trait variation of a genotype is specified by the phenotypic variance, which includes genetic and environmental components and GEI. Genetic variation is the prerequisite of selection of desirable genotypes and consequently plant improvement. GEI is of major importance to the plant breeders in developing improved cultivars. Plant breeders need to determine to what degree the desired traits are heritable and to what extent they are influenced by the environment.

The influence of GEI on crops was well recognized by plant breeders. It has been pointed out that many low-input agriculture areas of the world have not benefitted from yield increases seen elsewhere, because breeding is mainly conducted in the presence of high inputs (Ceccarelli, 1996). GEI influences the performance of crop cultivars under the environments with different external inputs especially nitrogen fertilizers. Breeding selection could be conducted under the target level of inputs, in other words, breeding for target population of environments. However few breeding selection is conducted in sub-optimal or stress conditions. Precisely what selection procedures are appropriate is still in debate.

Investigation of the effect of GEI on the GY of rainfed lowland rice has been conducted in few Asian countries (Cooper and Somrith, 1997; Henderson et al., 1996; Wade et al., 1999). Rice lines used in these studies were selected to represent genetic diversity adapted to the rainfed ecosystems. These studies indicated large and complex GEI was one of the key factors that influence the slow progress in rice genetic improvement for rainfed lowland ecosystems. The presence of GEI had important implications for breeding strategies that aim to improve either broad or specific adaptation or some combination of both components of adaptation for rainfed lowland rice. Alternatively, where some aspects of the GEI are repeatable, GEI information could be utilized to select for components of specific adaptation to the relevant target environments. However, the relative importance of genotype, environment and GEI on GY is less studied for irrigated rice, although evaluation of candidate varieties for yield stability through multi-environment trials (MET) before release and registration is mandatory in many countries.

1.4 Association mapping

Association mapping (AM), also known as linkage disequilibrium (LD) mapping or association study, is one of the several techniques to identify marker-trait associations and has been used extensively in human and animal genetic studies (Dewan et al., 2006; Karlsson et al., 2007). It has also been proven to be an effective approach to dissect complex traits in

important crop species, e.g. maize, rice, barley, durum wheat, common wheat, sorghum, sugarcane, sugar beet and soybean (Abdurakhmonov and Abdukarimov, 2008). The principle of AM is based on LD, or non-random association of alleles at adjacent loci within a population. One important advantage of AM over traditional biparental QTL mapping is that AM can be conducted directly on relevant breeding materials, thus permitting direct inference from data analysis to the breeding program. Furthermore, phenotypic variation is observed for most traits of interest and marker polymorphism is higher than in biparental populations (Buckler and Thornsberry, 2002; Jannink and Walsh, 2002; Yu and Buckler, 2006; Zhu et al., 2008). However, alleles at low frequency in an association panel are more difficult to assess (Myles et al., 2009) and the presence of population structure can result in spurious associations between a phenotype and markers that are not linked to any causative loci (Lander and Schork, 1994; Platt et al., 2010). Spurious associations can be reduced by taking population structure into account in a mixed linear model (MLM) analysis (Pritchard et al., 2000a; Sillanpää, 2011; Yu and Buckler, 2006).

1.5 Justification

Better understanding GEI for GY of irrigated lowland rice will provide useful information for genetic improvement and will help to identify superior and stable alleles/genotypes across multiple environments. It also helps breeders to employ appropriate breeding and selection strategies in practical breeding programs. In order to incorporate the desirable traits in a cultivar, such as improved GY or related traits, breeders have to test the performance of these traits in multiple environments to evaluate their stability. MET of a population of advanced breeding lines or cultivars will also help to identify the target environment groups where the genotypes perform similarly. So the advanced breeding materials can be exchanged or released as cultivars directly among those environments.

Why use AM in plant breeding? Modern rice breeding has made great progress in improving productivity in the last about 50 years, however yield increases of crops including rice (Nguyen and Ferrero, 2006) have been very slow in the last two decades, partly due to the lack of genetic diversity. To enlarge the genetic diversity in modern rice breeding populations thus improve yield potential and stability, new allelic source for important yield traits need to be explored. Numerous QTLs/genes for GY and related traits have been fine-mapped or cloned in biparental populations. Whether these well-characterized genes/QTLs are useful in a breeding population is not systemically studied. Validation of the effect of

genes/QTLs is prerequisite for utilization in breeding via MAS. Association study is useful tool to test the usefulness of known genes/QTLs in diverse genotypes.

Genome wide association studies (GWAS) has been proven to be an effective approach to dissect complex traits in important crop species including rice. In contrast to traditional QTL mapping, which was confined to bi-parental populations and can only capture allelic variation present in the two parents, association mapping sidesteps the need for bi-parental mapping populations and can be used to associate genotype with traits in breeding or other populations which have been extensively phenotyped. A large number of QTLs or genes for yield, yield component and grain quality have been finely mapped or cloned. Association analysis study using the population of mainly elite breeding lines will determine if it has sufficient power to detect these QTLs of known effects. The usefulness of these identified genes/QTLs in a breeding program will be tested as well. To enlarge the genetic diversity in modern rice breeding populations and improve yield potential and stability, new allelic sources for important yield traits need to also be explored. Candidate gene based association mapping method can be used to validate the QTLs or genes which have been identified by biparental populations QTL mapping. Genome wide association mapping has been proven to be an effective approach to dissect complex traits in important crop species including rice. Traditional QTL mapping was confined to bi-parental populations and only allelic variation present in the two parents can be captured. Association mapping sidesteps the need for bi-parental mapping populations and can be used to associate genotype with traits in breeding or other populations which have been extensively phenotyped.

1.6 Objectives

The present study was aiming at:

1. To get a well understanding of the GEI for GY of irrigated lowland rice by investigating the effects of genotype, environment including location, season and nitrogen rate on GY and related traits;
2. To test the usefulness of 39 fine-mapped or cloned genes/QTLs for GY and yield related traits through association analysis in an *indica* breeding population representing the genetic diversity of breeding gene pools for irrigated ecosystems at IRRI.
3. To identify associations between SNP markers and GY or yield related traits in an *indica* rice breeding population genotyped with GBS approach through GWAS.

Chapter 2 Literature review

2.1 Importance of rice

Rice is the most important food crop and a staple food of more than half of the world's population. More than half world population depend on rice for more than 20% of their daily calorie intake (GRiSP, 2010). Rice is grown under diverse environmental conditions. There are four major ecosystems: irrigated, rainfed lowland, upland and flood prone land (Khush, 1984). The total harvested area of rice was approximately 158 million ha and paddy rice was more than 700 million tons annually in 2009 (Seck et al., 2012). Asia produced about 90% of the rice in the world (nearly 640 million tons), with China as the leading producer. Rice consumption can exceed 100 kg per capita annually in many Asian and African countries. The world's population is expected to increase by about two billions in the next two decades and half of the increase will be in Asia (Gregory et al., 2000). It is estimated that the global rice production must reach 800 million tons in 2025 (the world paddy production in 2014 was 744.4 million tons) (FAOSTAT, 2014) to meet the demand for rice consumption. This challenge has to be overcome by developing new rice varieties with high-yielding potential and tolerance to biotic and abiotic stresses. The current average yields is about 5 ton/ha, which is far behind its potential of 10 tons ha^{-1} (Khush and Jena, 2009). The difference between potential yield and harvest yield is mainly due to genotype-by-environment interactions (GEI).

2.1.1 Importance of irrigated rice

Irrigated rice is grown in fields where irrigation systems supply and control the depth of water (5 to 25 cm). Almost all the rice in high-yield countries in Asia such as Korea, Japan and China is grown under irrigation (Chandler, 1979; Swain et al., 2005). Most of the rice grown under such ecosystems is transplanted and fertilizer is usually applied. Therefore the average yield is higher than those obtained from other rice ecosystems. Worldwide, irrigated rice accounted for 55% of the global harvested area and contributed 75% of the annual rice production (Dobermann et al., 2004). The average yield of irrigated lowland rice is about 5.4 tons ha^{-1} (GRiSP, 2010), with that in Indonesia, Korea, India and China being 6.3 tons ha^{-1}, 9.1 tons ha^{-1}, 3.6 tons ha^{-1} and 5.9 tons ha^{-1}, respectively. In temperate regions where a single irrigated rice crop is grown per year, the yield can reach 8-10 tons ha^{-1} (GRiSP, 2010). Irrigated rice has an important role in global food security with irrigated lowland rice being

cultured in around 85-90 million ha worldwide and provides 75% of the world's rice production (Seck et al., 2012). In Asia, the irrigated ecosystem constitutes 56% of the total rice area (Swain et al., 2005).

2.1.2 Yield barrier of irrigated rice

The Green Revolution has greatly increased rice production around the world. However, the grain potential of modern rice cultivars has remained stagnant for the past several decades (Nguyen and Ferrero, 2006). Studies had showed that there has been no progress in potential yield using side-by-side trials of inbred *indica* rice cultivars released in the Philippines at different times after 1968 (Peng et al., 1999, 2000, 2010). Rather, breeder's efforts have served to maintain a yield potential equivalent to that of the earliest cultivar IR8, released in 1966, and which currently yields a good deal less than at the time of its release in the face of some evolving (and, as yet, unidentified) biotic or abiotic challenges.

2.2 Commonly used breeding methods in rice

2.2.1 Pedigree breeding

Pedigree method is the major method used to develop improved rice varieties. The pedigree method is a selection protocol utilized during the inbreeding of populations in the development of homozygous lines. Superior single plants are selected in successive generations and a record is maintained of the parent-progeny relationships (which may extend to grand-parent, great-grandparent, or more ancestral generations). Selections are most frequently based on visual evaluations of high heritability traits, but any selection protocol, such as laboratory evaluation of plant chemical composition can be incorporated, if desired. It is estimated that more than 85% of the released rice varieties are developed using pedigree method (Guimarães, 2009). When there are possibilities to carry out multiple generations per year (e.g. off-season nurseries) this method is combined with modified bulk or even single-seed descent (SSD) to accelerate the process of having pure lines for agronomic trait evaluation.

2.2.2 Hybrid breeding

The development of hybrid rice varieties to utilize hybrid vigour and heterosis is one of the most successful methods for increasing rice yield. It has leaped rice yield to a new level in China. Hybrid rice usually produce 10%-15% higher yield than the inbred varieties. Hybrid rice occupies about 45% of the rice land in China (Khan et al., 2015). Two well-established systems for approaches for hybrid rice production are three-line method or CMS (cytoplasmic male sterility) system and two-line method or PTGMS (photo/temperature sensitive genic male sterility) system. The so-called one-line method that uses apomixis for utilizing heterosis is actively studied but has not been used in breeding practice (Cao and Zhan, 2014).

2.2.3 Ideotype breeding

IRRI postulated that the yield stagnation observed in the *indica* varieties in the tropics might be the result of the plant type of these varieties. They produce a large number of unproductive tillers and have excessive leaf area that may cause mutual shading and a reduction in canopy photosynthesis and sink size. A new plant type (NPT) was designed in 1989 based on previous physiological studies, breeders' experiences and the results of simulation modelling (Peng et al., 1999). The NPT has low tillering capacity (3-4 tillers when direct seeded); few unproductive tillers; 200-250 grains per panicle; a plant height (PH) of 90-100 cm; thick and sturdy stems; leaves that are thick, dark green, and erect; a vigorous root system; 100-130 days' growth duration; and increased harvest index (Peng et al., 1994). The first generation NPT lines were developed using tropical *japonica*. The NPT lines had large panicles, few unproductive tillers, and lodging resistance. However, grain yield (GY) was disappointing due to the low biomass production and poor grain filling. Reduced tillering capacity might contribute to low biomass production because the crop growth rate during the vegetative stage of NPT lines was lower. The first generation NPT lines are also susceptible to diseases and insects and have poor grain quality. Therefore, they could not be released for rice production in farmers' fields. However, the first generation NPT lines have been sent to more than 90 countries for evaluation. This valuable germplasm has been used as genetic materials in rice breeding programs worldwide. The second generation NPT lines were developed by crossing first generation tropical *japonica* NPT lines with elite *indica* parents. The following are the target traits: 330 panicles per m^2 , 150 spikelets per panicle, 80% grain filling, 25 mg grain weight (oven-dry), 22 tons ha^{-1} aboveground total biomass (at 14% moisture content), and 50% harvest index (Peng and Khush, 2003). One second generation NPT line, IR77186-

122-2-2-3, was released under the name of NSIC Rc158 in the Philippines in 2007. Three NPT lines have been released in China and one in Indonesia. However, there appears to be little or no realized gain in yield potential in these NPT lines with respect to earlier inbred cultivars.

Stimulated by IRRI's NPT breeding program, China established a nationwide mega project on the development of "super" rice in 1996 (Cheng et al., 2007) by developing inbred and hybrid varieties with ideotype. The plant type of China's "super" hybrid rice has many similarities with IRRI's NPT design. Both emphasize large and heavy panicles with reduced tillering capacity and improved lodging resistance. It was expected that harvest index could be improved with increased sink size and few unproductive tillers. Other common traits are erect leaf canopy and slightly increased PH in order to increase biomass production. However, the plant type of China's "super" hybrid rice give great emphasis on the top three leaves and panicle position within the canopy. Length, angle, shape, thickness, and area of the top three leaves were quantitatively defined. Many super hybrid varieties with significantly higher yield have been developed in China. However, whether the success is due to selection for the defined ideotype is unknown, since no solid data from systematic comparisons between selection using different criteria (e.g. ideotype) were collected.

2.2.4 Use of major quantitative trait loci

A large number of genes/QTLs for various traits have been identified (see section 2.4.1), which opened the possibility of applying marker-assisted selection (MAS) to select for or against targeted alleles. MAS allows for the selection of genes that control traits of interest. This technique is particularly useful in phenotype screening, which is expensive, technically difficult, or even impossible when using conventional methods. The selection process can be independent from phenotype, which allows selection in off-season nurseries, making the technique more cost effective to grow for more generations per year. Another benefit of MAS is the sharp reduction of required population size because many lines can be discarded in earlier breeding generations after MAS. The efficiency and usefulness of MAS for traits of simple inheritance (i.e. qualitative traits controlled by one or a few genes) have been well proven in many crops, including rice (Collard and Mackill, 2008; Guo and Ye, 2014). The following basic MAS strategies are usually applied: (1) backcrossing favorable alleles into elite germplasm, i.e. marker-assisted-backcrossing (MABC); and (2) stacking multiple genes of different sources, i.e. marker-assisted gene pyramiding (MAGP). The success of MAS has

motivated rice breeders to search for QTLs for complex traits, which account for a large proportion of phenotypic variation (major QTLs). Interestingly, major QTLs have been often reported for many yield-related traits (Miura et al., 2011).

2.2.4.1 Marker-assisted backcrossing

In MABC, markers are used during repeated backcrossing to select the presence of the target gene (foreground selection), to select against donor genome contribution (background selection), and to reduce the introgressed segment size and consequently linkage drag (recombination selection).

A converted version of rice restorer line Minghui 63 was developed using MABC with foreground selection for the target gene and recombination selection for breaking the linkage drag and the background selection for recovery of the recurrent genome (Chen et al., 2000). Minghui 63, which is susceptible to rice bacterial blight, is a widely-used rice restorer line for hybrid production in China. The gene *Xa21,* which confers wide spectrum resistance to bacterial blight, was successfully introgressed from isogenic line IRBB21 into Minghui 63 via MAS. A total of 128 evenly distributed restriction fragment length polymorphism (RFLP) markers were used to recover the genetic background of Minghui 63. It took three generations of backcrosses and one generation of selfing to complete the entire scheme.

The most well-known application of MABC in rice may be the conversion of high-yielding mega varieties into submergence tolerance ones by transferring the cloned submergence tolerance gene *SUB1*. *SUB1* was transferred to many mega varieties, such as Swarna, Samba Mahsuri, IR64, Thadokkam 1 (TDK1), CR1009, and BR11. Evaluations for submergence tolerance in the greenhouse and farmers' fields confirmed that all lines with *SUB1* QTL showed significantly greater tolerance to complete submergence as compared with their original parents (Sarkar and Bhattacharjee, 2011; Sarkar et al., 2009; Septiningsih et al., 2008; Singh et al., 2009). A single major QTL for salinity tolerance (*Saltol*) was identified on the short arm of chromosome 1, which explains much of the salt tolerance variation in a segregating rice population (Bonilla et al., 2002). Molecular markers closely linked to *Saltol* were used to transfer the QTL to commercial varieties, including BR11, BRRI dhan 28, BRRI dhan 29, and IR64 (Guo and Ye, 2014). Fourteen QTL for yield under drought were identified using GY as a selection criterion (Kumar et al., 2014). Six of these (*qDTY1.1, qDTY2.2, qDTY3.1, qDTY3.2, qDTY6.1* and *qDTY12.1*) showed an effect against two or more high-yielding genetic backgrounds in both the lowland and upland ecosystem,

indicating their usefulness in increasing GY under drought conditions (Kumar et al., 2014). *qDTY12.1* has been successfully introgressed in the background of Vandana. Vandana-introgressed lines with *qDTY12.1* showed a yield advantage of 0.5 tons ha^{-1} over drought tolerant cultivar Vandana under drought conditions and have a yield similar to that of Vandana under normal irrigation situations (Swamy and Kumar, 2012). Other popular varieties, i.e. Swarna, IR64, Sabitri, TDK1, Anjali, and Sambha Mahsuri, are improved or being improved by the introgression of QTL for GY under drought conditions (Kumar et al., 2014).

2.2.4.2 Maker-assisted gene pyramiding

Gene pyramiding is a method aimed at assembling multiple desirable genes from multiple parents into a single genotype. The end product of a gene pyramiding program is a genotype with all of the target genes. As foreground selection in the context of MABC, markers are used to ensure the presence of the target QTLs. The development of pyramided lines with multiple resistance genes is a commonly adopted strategy in breeding for resistance to major rice diseases (Ye and Smith, 2010).

MAS was also used to pyramid genes of yield and related traits to improve GY in rice. *Gn1a* is a gene for cytokinin oxidase dehydrogenase (OsCKX2), which controls grain number and contributes to increase grain productivity in rice (Ashikari et al., 2005). *sd1* encoding bibberellin20 oxidase was a QTL for PH (Spielmeyer et al., 2002). QTL pyramiding to combine *Gn1a* and *Hd1* in one line generated lines exhibiting both beneficial traits (Ashikari et al., 2005). The pyramided line showed increased grain production (23%) and reduced PH (20%) compared with Koshihikai. Qian et al (2007) summarized the development of a collection of pyramided lines containing different numbers of QTLs (2 to 9) for tiller number at the China National Rice Research Institute. Pyramided lines with desired tilling were obtained for super rice breeding. Phenotypic and genotypic information were used to develop a genotype-phenotype database, which allows for predicting the outcomes of pyramiding. Two QTLs, *qSBN1*and *qPBN6*, which increases the number of secondary rachis branches and the number of primary rachis branches, respectively, were introgressed into Sasanishiki from a high-yielding *indica* cultivar, Habataki. The new lines produced 62% to 65% more spikelets per panicle than Sasanishiki but only 4% to 12% higher yield (Ohsumi et al., 2011). The lines with both QTLs (SBN1 +PBN6) produced more spikelets than those with only one of the QTLs, NIL(SBN1) or NIL(PBN6) (Ando et al., 2008). Zong et al (2012)

pyramided eight QTLs for spikelet number per panicle and 1000-grain weight into variety 93-11 by using four recombinant inbred line (RILs). Compared with the parent line, the pyramided lines showed more spikelets per panicle but no changes for 1000-grain weight.

NILs containing one or more target genes were produced by transferring the alleles at *qHD8*, *qHD7* and *qHD6.1* (for heading date) and *GS3* (for grain size) from variety 93-11 into Zhenshan 97 (Wang et al., 2012). The line containing *qHD8* + *GS3* showed higher yield potential, longer grains and a more suitable heading date (HD) than Zhenshan 97. The line with *qHD7* + *qHD6.1* + *GS3* could also increase yield and leaf size, but considerably decreased days to heading. Four major drought QTLs, i.e. *qDTY2.2*, *qDTY4.1*, *qDTY9.1* and *qDTY10.1* were introgressed into IR64 background. Introgressed lines with two or more QTLs displayed yield advantages of 1.2 to 2.0 t/hm^2 under drought conditions while the yield and quality traits were similar to IR64 under normal irrigated situations (Swamy and Kumar, 2012).

IRRI and the Chinese Academy of Agricultural Sciences (CAAS), Beijing, China carried out a large-scale marker-assisted gene pyramiding program to improve drought tolerance. Segregating F_2 populations were produced by crossing introgression lines containing un-related QTLs for drought tolerance. The F_2 populations were then screened under both drought and normal non-stress conditions to identify superior individuals with desirable QTL or QTL combinations and high yield potential. Ten F_2 populations developed by this method at IRRI were tested under drought conditions, and 560 drought-tolerant pyramiding lines were identified. CAAS developed a total of 56 F_2 populations and selected 1,207 drought-tolerant pyramiding lines (Li et al., 2005).

2.2.5 Genetic modification

Genetic modification (GM) is a very efficient method for introducing foreign genes into rice genome. Protocols for efficient rice transformation using bombardment or Agrobacterium have been well developed (Christou et al., 1991; Hiei et al., 1994).

Two GM rice varieties with herbicide resistance, LLRice60 and LLRice62, were approved in the United States in 2000. Subsequent approval of these and other types of herbicide-resistant GM rice occurred across Canada, Australia, Mexico, and Colombia. However, none of these approvals resulted in commercialization. Varieties containing the *Bt* genes (*cry*1Ab, 1Ac 1Aa, 2A, 1B, or a combination of these genes) for resistance against lepidopteran pests were developed in different laboratories (Breitler et al., 2004; High et al.,

2004; Ho et al., 2006). The first field testing of the *Bt* rice was conducted in China in 1998 (Shu et al., 2000; Ye et al., 2001). Systematic field trials on these GM rice have shown that using the GM rice can reduce the use of pesticide significantly compared with using the conventional varieties. In 2009, China granted biosafety approval to GM rice varieties *Bt* Huahui No. 1 and *Bt* Shanyou 63 with pest resistance (Li et al., 2015). It is expected that China will be the first country to commercially release insect resistant transgenic rice.

Rice has also been engineered to withstand different abiotic stress conditions, such as drought, heat, cold, salinity, and mineral deficiency. Abscisic acid (ABA) is a phytohormone which plays important roles in the regulation of seed dormancy and adaptation to abiotic stresses. OsPYL/RCARs were identified as functional ABA receptors regulating ABA-dependent gene expression in rice. The overexpression of OsPYL/RCAR5 in rice driven by the maize ubiquitin promoter enhances improved drought and salt stress tolerance in rice (Kim et al., 2014). However, the plant height was slightly reduced under paddy field conditions and the grain yield severely decreased. It is necessary to fine regulating the expression level of OsPYL/RCAR5 to avoid deleterious effects on agricultural traits.

Nutritional improvement of rice grains has been a hot area in rice genetic engineering. The most well-known example is the development of Golden Rice. Vitamin A deficiency is one of the most prevalent deficiency diseases in developing countries, affecting more than 4 million children each year, up to 500,000 of whom become partially or totally blind (Ye et al., 2000). The entire β-carotene biosynthesis pathway has been engineered into rice endosperm, which is known as Golden Rice (Ye et al., 2000). Since then, efforts have been made to introduce genes for provitamin A biosynthesis into commercial rice varieties, including the genes for phytoene synthase (*psy*) and lycopene β-cyclase (*β-lcy*) originated from the daffodil and the gene for phytoene desaturase (*crt1*) of bacterial origin. 'Golden Rice 2' in which maize *psy* replaced daffodil *psy* contained carotenoids up to 23-fold of Golden Rice (Paine et al., 2005). Similarly, rice has been engineered to increase iron content, which is a very useful supplement for children and women in developing countries. At IRRI, two genes, ferritin and transporter gene, were added to IR64. The ferritin coding for iron storage is from soybean. Rice has its own ferritin gene, but adding additional ferritin gene increases the plant's iron storage capacity (Oliva et al., 2014). The transporter gene from another rice variety allows iron in the leaf to be transported to the grain. IRRI is also developing iron-rich rice and drought tolerant rice by genetic modification. Rice has also been transgenically modified to increase quantities of various amino acids and improve starch biosynthesis and oil quality (Newell-mcgloughlin, 2008).

Sun et al. (2015) engineered rice that stores more sugar in its grains and stems by adding a gene (SUSIBA2) from barley that affects starch storage. The new plant emits as little as 1% of the methane. What's more, the new rice may also boost food security as it produces significantly higher yield per plant. In a three-year-long trial, the rice grew well and led to drops in paddy field methane emissions.

Much effort has been made in the past decade to engineer C4 rice for higher photosynthesis efficiency and nitrogen and water use efficiency (Leegood, 2013). The International C4 Rice Consortium (http://c4rice.irri.org/) was formed to transfer high-yield C4 metabolism to rice and altering rice leaf anatomy and morphology in order to make it comparable with the Kranz-type biochemistry. The main approaches used in this ambitious project include: (i) integrating of the genes typical for the C4 metabolism into rice to increase photosynthesis efficiency, e.g. the genes encoding phosphoenolpyruvate carboxylase in mesophyll cells (MCs) and enzymes from Calvin-Benson cycle in bundle sheath cells (BSCs), (ii) down-regulating endogenous genes in rice, such as encoding MC enzymes of the Calvin-Benson cycle and photorespiration, (iii) introducing C4 cell-type specific gene expression and protein accumulation in rice, including the identification of suitable regulatory elements to ensure the protein's compartmentalization between MCs and BSCs, and (iv) identifying C4 transporters which transfer metabolites between subcellular compartments and introducing corresponding genes into rice (Karki et al., 2013; Leegood, 2013). Significant progress has been made through the identification of gene promoters for compartmental gene expression (Wang et al., 2013), gene cloning for the main enzymes of the C4 metabolic pathway from maize and their transformation into rice (Kajala et al., 2011), and the determination of candidate transporters of intermediate metabolites between MCs and BSCs (Karki et al., 2013).

2.3 Molecular and genomics tools for rice breeding

2.3.1 High throughput genotyping

Benefiting from having the smallest genome of the major cereals and relative ease of genetic transformation, rice has become the model organism for the cereal crops. Rice is the second plant genome to be completely sequenced, after the model plant species *A. thaliana*. The whole genome sequencing has been completed for both *japonica* and *indica* (Eckardt, 2000; IRGSP, 2005). The complete genome sequence is publicly available (http://rice.plantbiology.msu.edu/). Recently, the genome sequence of African rice (*Oryza*

glaberrima) was completed (Wang et al., 2014). The next generation sequencing (NGS) technology makes it possible to conduct genome sequencing in large scale. The 3000 Rice Genomes Project alone has seqeunced 3000 rice accessions (Li et al., 2014; The 3000 Rice Genomes Project, 2014).

With sequence information, abundant genome-wide markers can be developed quickly. To allow quick genotyping of a large number of markers multiplexed fixed SNP array chips have been developed by carefully selecting informative, evenly spaced SNPs across the genome. Rice 1,536, 6K and 50K SNP chips were developed using Illumina platform (Chen et al., 2014; Zhao et al., 2010). The 44K SNP chip was produced on Affymetrix platform (Famoso et al., 2011; Zhao et al., 2011). The 6K chip was designed to target functional genes in addition to genome-wide loci (Yu et al., 2014). These SNP array chips provide an efficient and reliable genotyping tool for rice genomics-assisted breeding. With the rapid reduction of sequencing cost, sequencing-based genotyping is becoming popular due to the much reduced cost. Several protocols have been developed to obtain sets or subsets of genomic restriction fragments for NGS, such as restriction-associated DNA (RAD) (Baird et al., 2008), diversity arrays technology (DArT) (Sansaloni et al., 2011), complexity reduction of polymorphic sequences (CRoPS) (van Orsouw et al., 2007) and genotyping-by-sequencing (GBS) (Elshire et al., 2011). GBS, originally developed for crop plants, is a robust platform which allows SNP discovery and genotyping individuals simultaneously. This method is based on the reduction of genome complexity with restriction enzyme and an efficient, user-friendly multiplexed library construction. Due to low coverage and high variance in sequencing, there is high amount of missing values in the discovered SNP marker. However several imputation methods are available (Marchini and Howie, 2010) and more in development. GBS has been successfully applied in biparental populations, multiparental populations and natural and breeding populations in rice (Begum et al., 2015; Spindel et al., 2013).

2.3.2 Sequence-based mapping

Compared to previous marker systems, NGS is very efficient for map-based gene discovery because it simultaneously performs SNP discovery, SNP validation, and SNP genotyping in a mapping or mutant population. Recently, a few approaches that combine NGS and bulk segregant analysis (BSA) have been developed for the identification of candidate genes associated with a phenotype of interest. Two DNA pools derived from individuals from the

phenotypic extremes of a segregating population are sequenced. MutMap is a method based on sequencing of pooled DNA samples from the phenotypic extremes of a segregating population derived from a cross between a mutant of interest and the progenitor wild type line (Abe et al., 2012). This strategy was successfully used to identify causal SNPs in a gene (*OsCAO1*) for the pale green leaf mutant in rice. The results were validated by transformation method. MutMap+ was developed to identify causal mutations by comparing SNP frequencies of bulked DNA of mutant and wild-type progeny of M3 generation derived from selfing of an M2 heterozygous line (Fekih et al., 2013). MutMap+ does not require artificial crossing between the wild-type parental line and mutants. This method is therefore suitable for identifying mutations that cause early development lethality, sterility, or generally hamper crossing. MutMap+ offers the same advantage as the classical MutMap protocol. Because it is based on selfing and it enables precise and robust phenotyping of minor effect traits. However, none of MutMap or MutMap+ can identify mutations in the genome regions missing from the reference (gaps) when the re-sequenced line displays significant structural variation from the reference genome. MutMap-Gap was developed, which involves delineating a candidate region harbouring a mutation of interest using MutMap, followed by *de novo* assembly, alignment, and identification of the mutation within genome gaps (Takagi et al., 2013b). MutMap-Gap has been successfully used to isolate the *Pii* gene, a gene for blast resistance, from the rice cultivar Hitomebore using mutant lines losing *Pii* function (Takagi et al., 2013b).

A QTL-seq method was developed to rapidly identify QTLs using progenies from crosses between genetically different varieties (Takagi et al., 2013a). In this method, two bulks of DNA are applied to whole genome sequencing, one (H-bulk) from 20-50 progenies showing higher values and the other (L-bulk) from 20-50 progenies with lower values for the phenotype of interest. QTL-seq was applied to rice RILs and F_2 populations to identify QTLs for partial resistance to the rice blast disease and seedling vigor (Takagi et al., 2013a). Yang et al (2013) employed NGS-assisted BSA and identified six QTLs for seedling cold tolerance in rice using a large F_3 population (10,800 individuals).

By resequencing of 132 RILs of variety *LYP9* and the parental lines, a high-resolution linkage map was constructed (Gao et al., 2013). Based on this high quality map, the genome sequences of the parental lines were significantly improved and 43 yield-associated QTLs were detected. Among those, *DTH8* and *LAX1* were detected as candidate genes for two QTLs, *qSN8* (for spikelet number) and *qSPB1* (for secondary panicle branch number),

respectively. A genetic complementation test demonstrated that *DTH8* indeed represents *qSN8* (Gao et al., 2013).

2.3.3 Large scale gene identification through transcriptome and TILLING

Microarrays and RNA-sequencing (RNA-Seq) have been used to study transcriptomes in multifarious developmental and stress conditions (Agarwal et al., 2014). Microarray could measure the expression levels of thousands of genes in a single experiment (Lipshutz et al., 1999) and to identify transcriptionally active regions (TARs) in the genome, which has been used for genome-wide polymorphism surveys and the identification of mutations (Chauhan et al., 2011; Oliveira et al., 2008; Singh et al., 2015; Wang et al., 2014). Differentially regulated genes have been selected as targets for functional validation. Transgenic plants with altered expression of the genes were generated, followed by their detailed analysis. RNA-Seq is becoming a more attractive method for whole-transcriptome studies (Van Verk et al., 2013), which extends the possibilities of transcriptome studies to the analysis of previously unidentified genes and of splice variants (Wang et al., 2009).

Conventional reverse genetics technologies, such as insertional mutagenesis with T-DNA, transposon/retrotransposon tagging or gene silencing using RNA interference (RNAi), have been used for rice functional genomics (Bolle et al., 2011; Hirochika, 2010; Wang et al., 2013). Large-scale genome sequencing has provided new possibilities for the application of conventional mutation techniques in both forward and reverse genetics strategies. Making knockout lines and using the targeting-induced local lesions in genomes (TILLING) method is possible. TILLING takes advantages of traditional mutagenesis, sequence availability and high-throughput screening for nucleotide polymorphisms in a targeted sequence. Eco-TILLING is a method that uses TILLING techniques to identify natural mutations in individuals (Wang et al., 2012). TILLING populations have been developed for several crop plants including rice (Rakshit et al., 2010; Till et al., 2007). To identify rare mutations in rice and wheat, Tsai and colleagues (2011) developed a new approach called “TILLING-by-Sequencing” in which target genes were amplified from pooled templates representing 768 individuals per experiment and then. Eco-TILLING, has also been used to identify a drought tolerance transcription factor in rice (Yu et al., 2012), and genes associated with salinity stress tolerance in rice (Negrão et al., 2013).

2.3.4 Targeted genome editing

Targeted genome editing involves the introduction of targeted DNA double-strand breaks (DSBs) using an engineered sequence-specific nucleases (SSN), stimulating cellular DNA repair by nonhomologous end joining (NHEJ) or homology directed repair (HDR) recombination mechanisms in different species. Different genome modifications can be achieved depending on the repair pathway and the availability of a repair template. Currently, four types of engineered nucleases are used for genome editing: engineered homing endonucleases/meganucleases (EMNs) (Silva et al., 2011), zinc finger nucleases (ZFNs) (Townsend et al., 2009), transcription activator-like effector nucleases (TALENs) (Cermak et al., 2011), and CRISPR (clustered regularly interspaced short palindromic repeats)/Cas 9 (Cong et al., 2013; Mali et al., 2013; Sander and Joung, 2014).

Site-directed mutagenesis can be achieved via NHEJ-mediated. HDR could use a DNA template to replace the DNA sequence at the break point, sebsequently cause directed gene knock-in/correction at specific locations in the genome. HDR works in late S/G2 phases in the cell cycle whereas NHEJ functions throughout the entire cell cycle. Thereby, NHEJ is the major DSB repair pathway in eukaryotes. These SSN effects generate targeted genome modifications including mutations, insertions, replacements and chromosome rearrangements (Kole et al., 2015).

SSN also allow targeted genes stacking, i.e. the addition of several genes in close vicinity to an existing transgenic locus. This makes it possible to add multiple traits into crops with a low risk of segregation, which is difficult to achieve using traditional breeding or even conventional transformation method (Ainley et al., 2013). In cotton, two herbicide tolerance genes were integrated into a pre-existing insect control locus following targeted cleavage and homology-directed repair using a designed meganuclease (D'Halluin et al., 2013). Similarly, two herbicide resistance genes were integrated into a pre-integrated ‘trait landing pad’ comprising a zinc finger binding site flanked by sequences homologous to an incoming donor DNA allowed for the sequential in maize (Ainley et al., 2013). A method has been proposed to stack involving cassette exchange of selectable marker genes in maize which preferentially select potential targeted events over multiple cycles of sequential transformation (Kumar et al., 2015). This method involves a target locus containing a first set of genes and a ‘trait landing pad’ between the first set of genes and a first selectable marker gene. The first set of genes created by either targeted integration into a ‘safe harbour’ locus or random integration. A unique 500-1000 bp 3’ intron sequence was included in the promoter which drives the

selectable marker gene in the target locus. The unique 3' intron sequence was flanked by the same nuclease recognition sequence as found in the 'trait landing pad'. Following nuclease-mediated cleavage in the target locus, the 3' intron and 'trait landing pad' provide homology with an incoming donor DNA containing second selectable marker gene without promoter and a second 'trait landing pad' for future targeting. Similarly, a second set of genes can be introduced into the target locus using selection and the first selectable marker is removed, thereby becoming available for future use. In principle, this process could be repeated over multiple cycles of targeted transformation with the same selectable marker genes and nuclease recognition sequences being recycled with each round of gene stacking (Petolino and Kumar, 2016).

Designed nucleases were used to excise stably integrated transgenic sequences. Large rice chromosomal sequences (>100 kb) containing a cluster of five phytoalexin biosynthetic genes were deleted using nucleases designed to cleave endogenous genomic loci (Zhou et al., 2014). Unlike recombinases which require the pre-integration of recognition sites flanking the sequence targeted for deletion, the ability of designed nucleases to cleave virtually any DNA sequence provides a greater degree of flexibility with respect to transgene deletion options.

Gene editing using TALENs (Li et al., 2012; Shan et al., 2013a) and Cas9/sgRNA (Feng et al., 2013; Jiang et al., 2013; Shan et al., 2013b) has been applied in rice. The ability to obtain biallelic gene modifications in a single generation (Feng et al., 2013; Xu et al., 2014; Zhou et al., 2014), the opportunity to delete large chromosomal segments (Zhou et al., 2014) and the availability of replacing gene through homologous recombination (Feng et al., 2013) coupled with positive/negative selection schemes (Shimatani et al., 2015) make it a useful tool for rice to generate genetically modified plants for basic research of monocot plant growth. In addition, the ability of knocking out several target genes simultaneously or modification (Endo et al., 2015; Xie et al., 2015) should greatly accelerate rice research and the development of improved rice varieties.

Rice plants with resistance to bacterial blight was developed via TALEN-directed disruption of a natural TAL effector binding element (EBE) sequence in the promoter of the disease susceptibility gene, *Os11N3* (also called *OsSWEET14*) (Li et al., 2012). Several resistant plants without the TALEN genes that had been inserted into the T_0 generation were selected by testing the progeny from crosses. Stable disease resistance was observed in T_1 and T_2 generation plants. The coding sequence of *OsSWEET13*, which is a bacterial blight disease susceptibility gene was modified using CRISPR/Cas9 technology and converted to a null

mutant for disease resistance (Zhou et al., 2015). A nonfragrant rice variety was converted into a fragrant variety by knocking out the OsBADH2 gene encoding betaine aldehyde dehydrogenase (BADH2) using TALENs (Shan et al., 2015). Likewise, inactivating the *CAO1* gene significantly changed the branching pattern of rice (Miao et al., 2013).

2.3.5 Genomic selection

A new MAS method, named genomic selection (GS) or genome-wide selection, has been developed to exploit all genes affecting a trait of interest using high density genome-wide markers (Meuwissen et al., 2001). The key features of this method are that markers covering the whole genome are used so that potentially all the genetic variance is explained by the markers. GS is a three-step process: 1) prediction model training and validation, 2) breeding value prediction of selection candidates, and 3) selection based on these predictions. In GS model training, a training population consisting of germplasm having both phenotypic and genome-wide marker data is used to estimate marker effects. The marker effect estimates and the marker data of the selection candidates are used to calculate genomic estimated breeding values (GEBVs) for an individual by summing up all marker effects included in the model. Selection is then imposed on the selection candidates using GEBVs as the selection criterion. Thus, GS attempts to capture the total genetic variance with genome-wide marker coverage and effect estimates. GS has been high-lighted as a promising MAS method for quantitative traits (Desta and Ortiz, 2014; Nakaya and Isobe, 2012).

Different statistical methods have been proposed to achieve higher prediction accuracy, which is defined as the correlation between the GEBVs and true breeding values. The majority of the methods are related to mixed model methodology considering QTL effects as random and systematic environmental effects as fixed. These methods could be clustered into two groups according to the assumptions about the statistical distribution of the marker effects (Xu and Hu, 2010). The first group assumes that all markers have some effect on the trait of interest and that each marker effect has equal variance. Ridge regression best linear unbiased prediction (RRBLUP) is a typical method based on this assumption. The second group assumes that marker effects come from different statistical distributions, which include BayesA, BayesB, Bayesian shrinkage and several others. RRBLUP is theoretically equivalent to the most the genomic best linear unbiased prediction (GBLUP). GBLUP utilizes genomic relationships to estimate the GEBV of a genotype (individual). For this purpose, a genomic relationship matrix is used, estimated from DNA marker information.

The matrix defines the covariance between individuals based on observed similarity at the genomic level, rather than on expected similarity based on pedigree, so that more accurate predictions of GEBV can be made. In a Bayesian framework, effect of a marker is represented by distribution of a random variable that is determined by its prior distribution according to some assumptions. The relative accuracy of these methods depends on the strength of marker effects. Where markers are in high LD with a few large effect QTLs that capture most of the genetic variance, i.e. strong marker effects, Bayesian methods are most accurate, whereas the RR-BLUP method is most accurate when many markers have small effects. Software tools implementing most of the commonly used genomic selection models are available.

In rice, empirical studies investigating GS accuracy have been reported. Spindel et al. (2015) used a population of 363 elite breeding lines from the IRRI irrigated rice breeding program genotyped with 73,147 high quality SNP markers using GBS to study the effects of training population size and composition, statistical methods, and number of markers on selection accuracy for three traits with different heritability (GY, PH and flowering time (FT)). For all three traits, GS models had higher prediction accuracy than that using multiple linear regression (MLR) and pedigree records alone. Prediction accuracies ranged from 0.31 and 0.34 for GY and PH to 0.63 for FT. The marker density sufficient for GS in this collection of rice was that using one marker every 0.2 cM. RR-BLUP was the best performing statistical method for GY. For FT, the non-GS MLR method outperformed GS models. For PH, random forest produced the most consistently accurate GS models (Spindel et al., 2015). Accuracy of GS in a rice synthetic population developed for RS breeding was investigated by Grenier et al. (2015). The effect of the LD and MAF thresholds for selecting markers, the relative size of the training population and of the validation population, the tested trait and the genomic prediction models on the accuracy of GEBVs was investigated. Significant differences in accuracy were observed among the different levels of each factor investigated. Phenotypic traits had the biggest effect, and the size of the incidence matrix had the smallest (Grenier et al., 2015). The accuracy of GEBV across all cross validation experiments ranged from 0.12 to 0.54 with an average of 0.30. Xu et al. (2014) reported GBLUP method to predict hybrid rice performance. In this study, a population of 278 randomly selected hybrids derived from 210 RILs was used as a training sample and all 21,945 potential hybrids were predicted. The average yield of top 100 selections shows a 16% increase compared with the average yield of all potential hybrids.

The prediction accuracy can be improved by integrating genomic prediction with crop modelling in a single hierarchical model. A novel method integrating whole-genome markers with a phenological model was developed to predict the rice HD of untested genotypes in untested environments (Onogi et al., 2016). The method simultaneously infers the phenological model parameters and whole-genome marker effects on the parameters in a Bayesian framework. By cultivating backcross inbred lines of Koshihikari × Kasalath in nine environments and comparing this method with conventional genomic prediction, phenological modelling and two-step approach, the proposed and two-step methods tended to provide more accurate predictions than the conventional genomic prediction methods, particularly in environments where phenotypes from environments similar to the target environment were unavailable for training genomic prediction. The proposed method showed greater accuracy in prediction than the two-step methods in all cross-validation schemes tested.

2.4 Methods for genetic dissection of complex traits in crop plants

2.4.1 Linkage mapping

Linkage analysis, also known as linkage mapping or meiotic mapping, involves several steps in plants (Semagn et al., 2006), i.e. creating an appropriate mapping population derived a bi-parental cross, genotyping with molecular markers to generate a linkage map and phenotyping for the trait of interest.

The commonly used types of mapping populations in linkage analysis in rice include doubled haploid lines (DH) (Li et al., 2005; Ma et al., 2009; Martinez et al., 2005), RILs (Cao et al., 2010; Kobayashi and Tomita, 2008; Sun et al., 2009) or advanced backcross lines (Alam et al., 2011; Blair et al., 2006; Kunert et al., 2007; Narasimhamoorthy et al., 2006). The populations are usually derived from crosses between two inbred lines with highly contrasting phenotype of the target traits. The advantage with DH or RIL is that a large amount of seeds can be produced and therefore larger experiments at multiple environments and years can be performed. The power of QTL detection in a bi-parental population and the accuracy of parameter estimates depend strongly on the choice of the two parental lines. Thus, the QTL detected in bi-parental populations only represents a part of the genetic architecture of the trait and only two allele effects are estimated.

Various methods have been used for linkage mapping in crops. Regression methods are easier and computationally faster, which is the same as analysis of variance at the marker

loci (Soller and Brody, 1976). It can be done using stand software like SAS or ASREML. In classical QTL analysis using interval mapping, the LOD-score profile is constructed over different genomic positions and the highest LOD (logarithm of the odds, to the base 10) value in each chromosome (which is also higher than a predetermined significance threshold) is taken as evidence of putative QTL position (Lander and Botstein, 1989). Composite interval mapping (CIM) has received much attention and has been widely applied in QTL mapping studies (Zeng, 1993). This method utilizes a pair of markers to locate the QTL position and at the same time other markers as covariates to control the genetic background. These markers outside of the target interval co-factors serve as proxies for other QTLs to increase the resolution of interval mapping, by accounting for linked QTLs and reducing the residual variation. The major problem using CIM to map QTL is that it is difficult to choose the suitable markers to serve as covariates. To circumvent the weaknesses of CIM, modifications of CIM was developed, e.g. multiple interval mapping (MIM) (Kao et al., 1999) and inclusive CIM (ICIM) (Li et al., 2008), which allow for simultaneous estimation of additive and epistatic effects. Furthermore, several other approaches are available, including Bayesian methods (Hoeschele and VanRaden, 1993; Satagopan et al., 1996; Sillanpa and Arjas, 1999), the use of a genetic algorithm (Carlborg et al., 2000) and mixed model analysis of QTL (Xu and Yi, 2000).

Many factors affect the accuracy and the power of QTL detection of linkage mapping. These include: 1) number of genes controlling the traits and their genome positions, 2) distribution of gene effects and existence of gene interactions, 3) heritability of the trait of interest, 4) number of genes segregating in mapping populations (natural and experimental populations), 5) population type and size, 6) marker density and coverage and 7) statistical methodology and significance level used (Asins, 2002; Kearsey and Farquhar, 1998; Kearsey and Pooni, 1996). The more QTLs there are in the population, the smaller their individual contribution to the phenotypic variation and the more difficult to detect. The reliability depends on the heritability of the individual QTL. The number of markers is important for conducting a QTL analysis. Marker density is a function of detection power within a certain density range (Darvasi et al., 1993; Frisch et al., 1999; Piepho, 2001). Increasing marker density beyond 10 cM had little influence on the power of QTL detection and the standard errors of genetic effect estimates (Piepho, 2001). The sensitivity of QTL detection decreased significantly as the population size decreased (Li et al., 2006). Relatively greater population sizes are particularly important for traits with low heritability and QTL with small effects to the traits. In determining how many individuals to be included in the mapping population and

number of genotyped markers, it is need to consider the QTL detection power to detect and the precision of QTL localization. Increasing the markers in the region of an inferred QTL may not necessarily improve the precision of QTL localization unless there are many individuals or the QTL has large effect on the trait (Broman, 2001). The precision of QTL mapping for an experimental cross can be increased by the use of a genotypically selected sample of individuals rather than a random sample of the same size (Xu et al., 2005).

2.4.2 Association mapping

Association mapping (AM), also known as linkage disequilibrium (LD) mapping or association study, has been extensively used in human and animal genetic studies (Dewan et al., 2006; Karlsson et al., 2007). The principle of AM is based on LD, or non-random association of alleles at adjacent loci within a population. It takes advantage of historical and evolutionary recombination events occurred at population level and achieves a relatively high resolution. AM has three advantages over traditional linkage mapping: increased mapping resolution, greater allelic diversity and less research time (Yu and Buckler, 2006). Since AM can be conducted directly on relevant breeding materials, it thus permitting direct inference from data analysis to the breeding program. Furthermore, phenotypic variation observed for most traits of interest and marker polymorphism is higher than in biparental populations (Buckler and Thornsberry, 2002; Jannink et al., 2001; Yu and Buckler, 2006; Zhu et al., 2008). However, alleles at low frequency in an association panel are more difficult to assess (Myles et al., 2009) and the presence of population structure can result in spurious associations between a phenotype and markers that are not linked to any causative loci (Lander and Schork, 1994; Platt et al., 2010). Spurious associations can be reduced by taking population structure into account in analysis (Pritchard et al., 2000a; Sillanpää, 2011; Yu and Buckler, 2006). Candidate gene-based association mapping and GWAS are the two major forms of AM. The advantages and disadvantages of the two AM methods and linkage mapping were briefly listed in Table 2.1.

2.4.2.1 Linkage and linkage disequilibrium

Linkage equilibrium (LE) and LD are used in population genetics to describe linkage relationships of alleles at different loci in a population (Abdurakhmonov and Abdukarimov, 2008). LE is a random association of alleles at different loci and equals the product of allele frequencies within haplotypes, meaning that combinations of alleles at each locus in the

haplotypes (combination of alleles) occur at random. Thus the frequencies of each combination have equal values in a population. LD, also known as gametic phase disequilibrium (GPD), is defined as a non-random association of alleles at separate loci located on the same chromosome (Mackay and Powell, 2007). Linkage refers to the correlated inheritance of loci through the physical connection on a chromosome, whereas LD refers to the correlation between alleles in a population (Flint-Garcia et al., 2003). Although close linkage between alleles on the same chromosome generally translate into high LD, significant LD may also exist between distant loci, and even between loci located on different chromosomes, which is the result of other forces such as selection, mutation, mating system, or population structure (Soto-Cerda and Cloutier, 2012). In other words, two markers in LD indicate a non-random association between alleles, but do not necessarily correlate/associate with a particular phenotypic trait, whereas association implies a statistical significance and refers to the covariance of a marker and a particular trait.

LD is caused by a variety of mechanisms, some of which may work simultaneously (Jannink and Walsh 2002). Populations expanding from a small number of founders and through admixture are the two most common ones. Here are some of the common mechanisms:

1. Populations expanding from few founders, in which some haplotypes were lost and others, such as the haplotypes present in the founders will be higher than expected under equilibrium (Russell et al., 2000).

2. GPD arises in structured populations when allelic frequencies differ at two loci across subpopulations, irrespective of the linkage status of the loci. Admixed populations, formed by the union of previously separate populations into a single panmictic one, can be considered a case of a structured population where substructuring has recently ceased.

3. Negative GPD will occur between loci affecting a character in populations under stabilizing or directional selection as a result of the Bulmer effect.

4. Positive GPD will occur between loci affecting a character under disruptive selection.

5. When loci interact epistatically, haplotypes carrying the allelic combination favored by selection will also be at higher-than-expected frequencies.

Factors influencing LD can be classed into two groups: factors increasing LD and factors decreasing LD (Abdurakhmonov and Abdukarimov, 2008). Factors, such as new mutation, mating system (self-pollination), genetic isolation, population structure, relatedness (kinship), small founder population size or genetic drift, admixture, selection (natural,

artificial, and balancing), epistasis, and genomic rearrangements will increase LD, while high recombination and mutation rate, recurrent mutations, outcrossing, and gene conversions will decrease LD (Gupta et al., 2005). LD generally decays faster in cross-pollinated crops than in self-pollinated species, in which individuals are more inclined to be homozygous at a given locus, than in outcrossing species (Flint-Garcia et al., 2003). LD also decays faster in diverse populations compared to populations with narrow genetic background. Different genes and genomic regions in the same crop can exhibit different extent of LD decay (Mather et al., 2007). In the context of AM, the extent of LD decay determines the number of markers required in an AM analysis and the expected resolution of identified associations. A high level of LD requires reduced number of markers needed for AM, but it results in a lower resolution (coarse mapping). In contrast, a less extensive level of LD requires more markers to track a gene of interest, but the tracking can be done in high resolution (fine mapping).

2.4.2.2 Population structure detection and control in association mapping

Population structure and population admixture are the main factors to create LD between unlinked loci. This primarily happens due to the occurrence of distinct allele frequencies with different ancestry in an admixed or structured population. LD generated by selection, population structure, relatedness, and genetic drift might be theoretically useful for AM in specific situations and population groups that reduces number of markers needed for AM (Stich et al., 2006, 2005), but requires special attention to control factors affecting LD (e.g. population structure and relatedness) to perform unbiased population-based AM in plants (Liu and Muse, 2005; Pritchard et al., 2000a).

In a given association panel, the population structure can be detected with various statistical methods using random markers distributing through whole genome (Pritchard et al., 2000a, 2000b). These statistical methods could be clustered into two major categories: parametric and non-parametric approaches. STRUCTURE and L-POP are parametric approaches, which assume Hardy-Weinberg equilibrium (HWE) and LE among loci in individuals of the association panel. STRUCTURE, a model based Bayesian method, uses a Markov Chain Monte Carlo (MCMC) algorithm base on the Gibbs sampler algorithm (Pritchard et al., 2000a), and L-POP uses the Expectation-Maximization (EM) algorithm (Purcell and Sham, 2004). Another model-based approach, MCLUST, was reported to predict population structure without genetic assumptions (Fraley and Raftery, 2007). Non-parametric methods use a two-stage design, which are dimension reduction and statistical clustering

methods to separate individuals (Liu and Zhao, 2006). Dimension reduction could be achieved by calculating the pair-wise distances (Bowcock et al., 1994; Mountain and Cavalli-sforza, 1997; Shriver et al., 2004) or singular value decomposition (SVD) (Liu and Zhao, 2006). The commonly used statistical clustering methods are principal coordinates analysis (PCoA) (Bauchet et al., 2007; Shriver et al., 2004) and multidimensional scaling (MDS) (Lao et al., 2006; Purcell et al., 2007). Clustering method, e.g. neighbour joining (NJ) (Bowcock et al., 1994; Mountain and Cavalli-sforza, 1997), K-means method (Liu and Zhao, 2006), could be used to cluster individuals. In contrast to parametric approaches, non-parametric approaches do not require assumptions about population models, nor do they estimate allele frequencies. They could be complementary to parametric approaches in population structure analysis. Especially for those situations where parametric model assumptions cannot be verified, or there is only a limited number of individuals from a single sub-population, non-parametric methods are more powerful for inference.

Three methods are used to control the population structure in AM: (1) genomic control, (2) use of a fixed model and (3) use of a mixed model. The method of genomic control is built on a Bayesian probability model and uses random markers to evaluate the overall structure effects on *P* values; which are then adjusted to account for the statistical inflation caused by the structure (Devlin and Roeder, 1999). The fixed model approaches use molecular markers to estimate the population structure. These estimates are integrated in the AM tests as covariates. Similarly, mixed model approaches use both fixed and random effects to control the population structure (Stich et al., 2008b; Yu et al., 2006; Zhao et al., 2007). Furthermore, in order to control spurious associations, rare alleles (frequency <5%) are usually treated as missing data or discarded for population structure, LD analysis and AM (Breseghello and Sorrells, 2006).

2.4.2.3 Candidate gene-based association mapping

A candidate gene is a gene, located in a chromosome region with known biological function, and is directly or indirectly involved in the expression of the investigated trait (Tabor et al., 2002; Zhu and Zhao, 2007). The candidate genes are selected based on prior knowledge from mutational analysis, linkage analysis, biochemical pathway, transcriptomics and other omics approaches (Tabor et al., 2002). It is straightforward to select candidate genes for relatively simple biochemical pathways (e.g. starch synthesis in maize) or well characterized pathways (e.g. FT in *Arabidopsis*) (Zhu et al., 2008). Association analysis can validate a candidate gene by evaluating the effects of the causative gene variants. Only markers for the candidate genes

are required to perform candidate gene-based association mapping analysis. Usually an additional independent set of random markers is also included for the purpose of population structure inference (Pritchard and Rosenberg, 1999). Candidate gene-based association mapping directly focuses on the association between the trait and variants in particular genes that have a priori biological support for being involved the trait expression. This focus means candidate gene studies ignore much of the genome, and might miss many causal regions (Witte, 2010).

Table 2.1 Comparison of linkage mapping and association mapping

	Linkage mapping	Candidate gene-based association mapping	GWAS
Main advantages	No population structure effects; Identification of rare alleles; Few genetic markers required	Allows fine mapping Relatively low costs	Allows untargeted fine mapping; Detection of common alleles
Main disadvantages	Limited genetic diversity; Not always possible to create crosses; Cannot distinguish between pleiotropic and physically close genes	Detailed functional knowledge of trait is required; No novel traits will be found	Confounding effects due to population structure; Will miss rare and weak effect alleles
General requirements	Small 'original population size', low number of genetic markers, many replicates needed; Generated mapping material, e.g.F_2 population, (AI-)RILs, MAGIC lines, NILs, HIFs, etc.	Large population size, small number of markers; Prior genetic and biochemical knowledge on trait of interest; Prior knowledge on LD, population structure and breeding system	Large population size, many genetic markers; Prior knowledge on LD, nucleotide polymorphism, breeding system and population structure

From Kloth et al., (2012). Trends in Plant Science, 17:311-319.

AI-RIL, advanced intercross-recombinant inbred line; HIFs, heterogeneous inbred family; LD, linkage disequilibrium; MAGIC, multiparent advanced generation intercross; NIL, near isogenic line; RIL, recombinant inbred line; SNPs, single nucleotide polymorphisms.

2.4.2.4 Genome-wide association study

Genome-wide association study (GWAS), also known as genome-wide association mapping, is a comprehensive approach to systematically search the genome for causal genetic variation for a given trait. Compared to the candidate gene-based association mapping, it requires testing of a high density of markers (depending on both genome size and LD decay) covering the whole genome within the LD decay of the species. To identify the associations between

the various complex traits and causal genetic variation, prior information regarding candidate genes is not required. When one is interested in finding all genomic regions that may be involved in controlling the trait of interest, GWAS would be an appropriate option. Since adopted in plant genetics, GWAS has been proven to be a powerful tool for identifying QTLs for complex traits in plants (Ogura and Busch, 2015). GWAS has been applied in the major cereal crop species, including rice, maize, barley, bread wheat, and oats. With the rapid development in NGS technology, the genotyping cost reduced dramatically and high density genome-wide marker are available. GWAS is now becoming more and more popular in plant genetics.

2.5 Genetic mapping of grain yield and related traits in rice

2.5.1 Well characterized genes/QTLs

GY is one of the most important traits in rice breeding. It is governed by multiple QTLs with minor effects. GY is determined by three component traits: panicle number per plant (PN), grain number per panicle (GN), and thousand grain weight (TGW). Multiple agronomic or morphological traits, such as HD, PH, tiller number per plant (TN), primary branch per panicle (PN), secondary branch per panicle (SB), spikelet number per panicle (SN), and seed setting ratio (SR) all contribute to GY indirectly. Some grain shape related traits; e.g. grain size, grain length (GL) and grain width (GW) also affect the rice GY to different extent. More than 20 QTLs affecting rice GY and components have been cloned, and some QTLs have been validated (Bai et al., 2010; Zuo and Li, 2014) These are listed in Table 2.2 and some important and well characterized genes/QTLs will be discussed.

2.5.1.1 Thousand grain weight

Grain size is a major determinant of TGW. The increase in grain size is also one of the main components of the domestication syndrome in cereals, which distinguishes the domesticated species from its wild ancestors (Brown et al., 2009; Fuller, 2007). In breeding practice, the grain size is usually assessed by TGW, which is positively correlated with several characters including grain length, GW and grain thickness (Evans, 1972; Xu et al., 2002). Grain size also affects rice quality.

A major QTL affecting grain length on chromosome 7 (*GL7*) was cloned (Wang et al., 2015). The copy number at the *GL7* locus contributes to grain size in rice. *GL7* encodes a

protein homologous to *Arabidopsis* LONGIFOLIA proteins, which regulate longitudinal cell elongation. An increase in grain length and improvement of grain appearance can be achieved by tandem duplication of a 17.1-kb segment at the *GL7* locus. It was suggested that pyramiding beneficial alleles of *GL7* and other yield- and quality-related genes may improve the breeding of elite rice varieties (Wang et al., 2015).

GS3, a major QTL for grain size, has been frequently detected in many studies across different genetic backgrounds and environments (Aluko et al., 2004; Huang et al., 1997; Redoña and Mackill, 1998; Thomson et al., 2003). The *GS3* locus explained 80-90% of the variation of TGW and GL in a random subpopulation from the BC_3F_2 progeny. The *GS3* locus also showed minor but significant effect on GW and thickness (Fan et al., 2006). *gw3.1*, a QTL for TGW, was mapped to a 93.8kb-region on chromosome 3 using a population of NILs developed from the cross between *Oryza sativa*, cv. Jefferson × *O. rufipogon* (IRGC105491) (Li et al., 2004). *GW2*, responsible for TGW, encodes a previously unknown RING-type protein with E3 ubiquitin ligase activity. A loss of *GW2* function increased cell numbers, resulting in a larger (wider) spikelet hull, and it accelerated the grain milk filling rate, resulting in enhanced TGW and GY (Song et al., 2007). Two QTL, *GW3* and *GW6* controlling TGW and shape, were mapped on chromosome 3 and 6, respectively (Guo et al., 2009). The QTL, *tgw2*, was identified in a backcross population (Yoon et al., 2006). *qTGW7*, was found on the short arm of chromosome 7 and cosegregated with RM22034 in 108 chromosome segment substitution lines (CSSLs) that contained the introgressed chromosomal segments covering 98.3% of the cultivar C418 genome (Bian et al., 2010). A pleiotropic QTL, *qGL7*, exhibiting effects on TGW, GL, GW, SN and grain thickness was mapped to a 258-kb region flanked by InDel marker RID711 and SSR marker RM6389 using a NILs population (Bai et al., 2010)

2.5.1.2 Panicle number per plant

A body of tillering mutants has been validated and several of the corresponding genes have been isolated (Table 2.2). *Monoculm 1* (*MOC1*) encodes a putative GRAS family protein, which is mainly expressed in the axillary buds, and functions to initiate axillary buds and promote their outgrowth (Li et al., 2003). The allele of a major QTL (*qPN1*) from Nipponbare decreased the PN. *qPN1* was mapped to a 34.4kb chromosome region on chromosome 1 (Zhu et al., 2011). The *OsTB* gene is the rice gene orthologous to *TEOSINTE BRANCHED 1* (*TB1*) involved in lateral branching in maize. *OsTB* encodes a putative

transcript factor containing a TCP domain (Cubas et al., 1999), negatively regulates lateral branching in rice via expression in axillary buds (Takeda et al., 2003). The rice *HIGH-TILLERING DWARF1* (*HTD1*) encoding an ortholog of *Arabidopsis* MAX3 is required for negative regulation of the outgrowth of axillary buds (Zou et al., 2006). *DWARF3* (*D3*) and *DWARF10* (*D10*), which are *Arabidopsis MAX2* and *RMS1/MAX4/DAD1* orthologs respectively, control lateral bud outgrowth in rice (Arite et al., 2007). All the evidence suggests that the shoot branching regulation mechanisms are highly conserved between monocots and dicots. A novel phytohormone, strigolactones, which regulates the MAX/RMS/D pathway in both dicots and monocots, has also been identified (Gomez-Roldan et al., 2008; Umehara et al., 2008). The rice *D27*, *HTD2* and *D88* play important roles in the MAX/RMS/D pathway (Gao et al., 2009; Lin et al., 2009; Liu et al., 2009). The QTL, *Ideal Plant Architecture 1* (*IPA1*)/*Wealthy Farmer's Panicle (WFP)*, was found to encode OsSPL14. High expression of *OsSPL14* in the reproductive stage promotes panicle branching and higher grain yield in rice (Miura et al. 2010).

2.5.1.3 Grain number per panicle or spikelet number per panicle

Total grain number per plant is the most important factor for increasing GY under field production conditions (Ashikari et al., 2005). SN is positively correlated with GN (Ranawake and Amarasinghe, 2014). *gpa7*, a QTL controlling GN, was mapped to a 35-kb region on the short arm of chromosome 7. This region contains five predicted genes (Tian et al., 2006). A major QTL controlling SN, *qSPP7*, was mapped to a 912.4kb region on chromosome 7, cosegregating with two markers, RM5436 and RM5499 (Xing et al., 2008). Interestingly, this region also affects GY per plant, TGW, TN, and SR. *qSPP7* was cloned and renamed as *Ghd7*, controlling GN, PH and HD (Xue et al., 2008). *Gn1a*, a QTL contributes to increasing GN, encodes cytokinin oxidase/dehydrogenase (*OsCKX2*) that degrades the phytohormone cytokinin (Ashikari et al., 2005). Enhanced GY can be achieved by reducing the expression of *OsCKX2*. *SPP1*, a major QTL controlling the SN, was mapped near *Gn1a* and *Gn1b* on chromosome 1 (Liu et al., 2009) but the relationship between them is not yet clear. *SPP3a* and *TGW3b*, *TGW3b* and *SPP3b*, controlling SN and TGW, were simultaneously mapped to two different loci on chromosome 3 (Liu et al., 2010). However, these loci showed less effect on GY due to the increased SN but decreased TGW and vice versa. One major QTL (*qSPP7*) and three minor QTLs (*qSPP1*, *qSPP2* and *qSPP3*) were identified on chromosome 7, 1, 2 and 3, respectively. The four QTLs were validated in the corresponding NILs (Zhang et al.,

2009). Two QTLs for SN, *qNSP6.1* and *qNSP6.2* showed effects on the GN and GY. *qNSP6.2* was cosegregated with the HD gene *Hd1* (Gong et al., 2010). *Ideal Plant Architecture 1* (*IPA1*)/*Wealthy Farmer's Panicle (WFP)* was associated with an increased grain number.

2.5.1.4 Other yield related traits

The architecture of the panicle directly determines GY. The architecture of the rice panicle is mainly shaped by the arrangement of branches, spikelets and panicle erectness. Several genes controlling panicle branches have been identified. *Dense and erect panicle 1* (*DEP1*) is essential dominant/semi-dominant regulators that determine rice panicle branches and thus affect the GY (Huang et al., 2009). The alleles, *dep1* and *ipa1* have been successfully used for the improvement of GY via MAS (Wang and Li, 2011). QTLs were also identified for PB and the SB (Guo and Hong, 2010). A gene for PB (*ABERRANT PANICLE ORGANIZATION 1, APO1*) was mapped on chromosome 5 using backcrossed lines. This gene controls both PB and GN, thus has significant effects on GY per plant (Terao et al., 2010).

Table 2.2 Fine-mapped or cloned genes/QTLs for grain yield or related traits in rice

QTL[a]	Trait	Chr[b]	Interval[c]	Flanking/Closest markers[d]	Reference
SCM2	culm strength	Chr6	15 kb	InDel02-InDel08	Ookawa et al., 2010
dep3	dense and erect panicle	Chr6L	NA[f]	RM30-S06110	Qiao et al., 2011
dep3	dense and erect panicle	Chr6L	73 kb	P21-P23	Qiao et al., 2011
DEP1	dense panicle, grain number and errect panicle	Chr9	82 kb	s2-s11-2	Huang et al., 2009
DWL1	dwarfism and withered leaf tip	Chr3 L	46 kb	HL921-HL944	Jiang et al., 2008
ep3	erect panicle	Chr2S	46.8 kb)	STS5803-5--STS5803-7	Piao et al., 2009
qFLL9	flag leaf length	Chr9	198 kb	RM24423-RM24434	Jiang et al., 2010
DTH8	flowering, plant height and yield potential	Chr8S	47 kb	Ind8-47-Ind8-15	Wei et al., 2010
GIF1	grain incomplete filling	Chr4	32 kb	CAPS-4-CAPS-8	Wang et al., 2008
GL7	grain length	Chr7	20.4 kb	CAPS1-210Q	Wang et al., 2015
qGL7	grain length	Chr7	258 kb	RID711-RM6389	Bai et al., 2010
qGL-1	grain length	Chr1S	437.5 kb	RM10390-RM1344	Yu et al., 2008
qGL7-2	grain length	Chr7	278 kb	Indel1-RM21945	Shao et al., 2010
Gn1a	grain number	Chr1S	6.3 kb	3A28-3A20	Ashikari et al., 2005
qGN-1	grain number	Chr4L	11.1 cM	nksrssr04-02-HvSSR04-40	Deshmukh et al., 2010
qGN-1	grain number	Chr4L	NA	RM3276 cosegregate with	Deshmukh et al., 2010
qGN4-1	grain number	Chr4L	0.78 kb	nksssr04-02-nksssr04-19	Deshmukh et al., 2010
Gn1a	grain number per panicle	Chr1	NA	Gn1a-M1-Gn1a-M2	Yan et al., 2009
Gnp4	grain number per panicle	Chr4L	10.7 kb	Y42-Y48	Zhang et al., 2011
Gnp4	grain number per panicle	Chr4L	93.2 kb	RM16874-RM16888	Zhang et al., 2011
NGP	grain number per panicle	Chr6S	NA	RM111 to RM19784	Gong et al., 2010
GS3	grain size	Chr3	NA	RGS1 (function marker)	Wang et al., 2011
OsSPL16	grain size, shape and quality	Chr8L	7.5 kb	RM502-PSM711	Wang et al., 2012
TGW3b	grain weight	Chr3	2.6 cM	RM15885-W3D16	Liu et al., 2010
tgw6	grain weight	Chr6	NA	C358	Ishimaru, 2003
gw3.1	grain weight	Chr3	93.8 kb	JL123-JL109	Li et al., 2004
tgw2	grain weight	Chr2	384 kb	RM12813-RM12836	Oh et al., 2010
Gw1-1	grain weight	Chr1	392.9 kb	RM10376-RM10398	Yu et al., 2008

Gw1-2	grain weight	Chr1	308.5 kb	RM10404-RM1344	Yu et al., 2008
GW3	grain weight	Chr3	122 kb	WGW16-WGW19	Guo et al., 2009
gw5	grain weight and length-width ratio	Chr5	49.7 kb	RMw530-RMw513	Wan et al., 2008
GW6	grain weight	Chr6	NA	RM7179-RM3187	Guo et al., 2009
tgw2	grain weight	Chr2	4.4Mb	RM7288-RM13104	Yoon et al., 2006
GS3	grain weight	Chr3	NA	GS3-PstI	Yan et al., 2009
GW2	grain weight	Chr2	NA	GW2-HpaI	Yan et al., 2009
gw8.1	grain weight	Chr8	306.4 kb	RM23201.CNR151-RM30000.CNR99	Xie et al., 2006
qTGW7	grain weight	Chr7	NA	RM22034	Bian et al., 2010
GS3	grain weight and length	Chr3	7.9 kb	GS63-SF19	Fan et al., 2006
qGW-7	grain weight	Chr7	~2.6 kb	S5-S6	Wang et al., 2015
GW2	GW and weight	Chr2	8.2 kb	W024-W004	Song et al., 2007
GW5	GW and weight	Chr5	21 kb	Cw5-Cw6	Weng et al., 2008
GY	grain yield per plant	Chr6S	NA	RM111-RM19784	Gong et al., 2010
qGY2–1	grain yield per plant	Chr2S	102.9 kb	RM279-SBG1	He et al., 2006
Ghd8	grain yield, heading date, and plant height	Chr8	20 kb	Pa6-Pa7	Yan et al., 2011
Ghd8	grain yield, heading date, and plant height	Chr8	70 kb	SEQ3-1-SEQ5-1	Yan et al., 2011
Hd3a	heading date	Chr6S	20 kb	25-5UL-CP59	Kojima et al., 2002
Hd9	heading date	Chr3S	NA	C721-R1468B	Lin et al., 2002
Ghd7	grain per panicle, heading date and plant height	Chr7	0.31 cM	RM5436-C39	Xue et al., 2008
IPA1	ideal plant architecture	Chr8L	78 kb	M4-M5	Jiao et al., 2010
OsTB1	lateral branching	Chr3	NA	C944	Takeda et al., 2003
gpa7	less grains per panicle	Chr7S	35 kb	3617(SSR)-ID52(InDel)	Tian et al., 2006
Ltn	low tillering	Chr8L	38.6 kb	ssr0649-23-ssr0649-1(3)	Fujita et al., 2011
MOC1	monoculm	Chr6L	20 kb	17-2-12-2	Li et al., 2003
MOC1	monoculm	Chr6L	3.4 cM	R1559-S1437	Li et al., 2003
Nop(t)	non-panicle	Chr6L	102 kb	M9-M10	Wu et al., 2009
PAP2	panicle development	Chr3	10 kb	CAPS2-CAPS3	Kobayashi et al., 2010
PAP2	panicle development	Chr3	NA	RM15937-RM15948	Kobayashi et al., 2010
qPN1	panicle number per plant	Chr1L	34.4 kb	S1-86-S1-109	Zhu et al., 2011
qPN1	panicle number per plant	Chr1L	34.4 kb	S1-86-S1-109	Zhu et al., 2011

Pnn1	panicle number per plant	Chr2	6.3 cM	2-S152,2-S173	Obara et al., 2004
qPH6-1	plant height	Chr6S	51.7 kb	Si2925-RM19417	Bao et al., 2009
qPH6-1	plant height	Chr6S	96.4 kb	RM3414-RM19417	Bao et al., 2009
qRL6.1	root length	Chr6L	337 kb	MID06024-MID06029	Obara et al., 2010
qSW5	seed width	Chr5	2.26 kb	MS40671-M16	Shomura et al., 2008
SP1	short panicle	Chr11	8 kb	M7-M8	Li et al., 2009
spd6	small panicle and dwarfness	Chr6	22.4 kb	Q5-JX6036	Shan et al., 2009
SPP1	spikelet number per panicle	Chr1	107 kb	YN27-YN34	Liu et al., 2009
NSP	spikelet number per panicle	Chr6S	NA	RM111 to RM19784	Gong et al., 2010
qSPP7	spikelet number per panicle	Chr7	912.4 kb	RM5436-RM5499	Xing et al., 2008
qSPP7	spikelet number per panicle	Chr7	NA	0.7 cM from MRG4436	Zhang et al., 2009
qSPP1	spikelet number per panicle	Chr1	NA	0.5 cM from MRG2746	Zhang et al., 2009
qSPP2	spikelet number per panicle	Chr2	NA	0.6 cM from MRG2762	Zhang et al., 2009
qSPP3	spikelet number per panicle	Chr3	NA	0.8 cM from RM49	Zhang et al., 2009
QSpp8	spikelet number per panicle	Chr8S	1.4 cM	RM310-RM126	Xing et al., 2008
SPP3b	spikelet number per panicle	Chr3	2.6 cM	RM15885-W3D16	Liu et al., 2010
qPDS3	spikelet on the primary branches	Chr3	1.6 cM	RM14820-RM14823	Tan et al., 2011
htd1	tillering and dwarfness	Chr4	30 kb	C2-D1	Zou et al., 2005
D88	tillering and dwarfness	Chr3S	14.5 kb	P3-2-P3-4	Gao et al., 2009
D10	tillering and dwarfness	Chr1	2.2 cM	RM1095-Rm3411	Arite et al., 2007
d27	tillering and plant height	Chr11L	18 kb	P3-P6	Lin et al., 2009
d27	tillering and plant height	Chr11L	3.0 cM	C189-RM206	Lin et al., 2009
WFP	wealthy farmer's panicle	Chr8	66 kb	dCAPS825-CAPS311	Miura et al., 2010
gw1	yield enhancing	Chr9	37.4 kb	RM24718.CNR111-RM30005.CNR142	Xie et al., 2008
gw9.1	yield-related	Chr9L	37.4 kb	RM24718.CNR111-RM30005.CNR142	Xie et al., 2008

[a] QTL, mapped for yield or yield component traits.

[b] Chr, chromosome which the QTL located on.

[c] Interval, the length between the two markers flanking the peak position of a QTL.

[d] Flanking/Closest markers, the genetic marker which flanking the QTL or the nearest marker.

[e] NA, Not available.

2.5.2 Association mapping in rice

2.5.2.1 Population structure of rice

Population structure and familial relatedness are two of the major confounding factors for GWAS (Zhang et al., 2010). In order to reduce the false positive results caused by population stratification, population structure should be carefully taken into account in association mapping. Population structure analysis using different analysis methods in diversity panels of different sizes has indicated the existence of two to eight subpopulations in rice (Agrama et al., 2007; Chakhonkaen et al., 2012; Huang et al., 2012, 2010; Nachimuthu et al., 2015; Zhang et al., 2009; Zhang et al., 2009; Zhao et al., 2011). The two major subgroups, e.g. *indica* and *japonica*, are resulted from the different adaptation behaviour of accessions to different ecological environment as *indica* and *japonica* accessions has independent evolution frame (Huang et al., 2010; Nachimuthu et al., 2015; Wu et al., 2015; Zhang et al., 2011). An European core collection of rice was classified two subpopulations as *japonica* and non-*japonica* accessions (Courtois et al., 2012). Using different collections and different methods, more than two subpopulations were also reported (Huang et al., 2012; Jin et al., 2010; Kumar et al., 2015; Wang et al., 2013; Zhang et al., 2009; Zhao et al., 2011).

Studies also have been extensively conducted to examine the population structure within *indica* rice panel only (Begum et al., 2015; Singh et al., 2013; Wang et al., 2014; Xie et al., 2012) and subpopulations were observed within *indica* subspecies. For examples, 375 India *indica* rice varieties were divided into five subgroups with 36 SSR markers while it was 15 subgroups using 36 SNPs (Singh et al., 2013). 1482 Chinese *indica* landraces were divided into three ecotypes, viz. early (*Ind.*E), late (*Ind.*L) and intermediate (*Ind.*M) and further into nine eco-geographical types (Zhang et al., 2013). A collection of 215 widely cultivated *indica* rice varieties developed in southern China and IRRI were clustered into two major subpopulations, in which IRRI varieties were closely grouped and separated clearly from the majority of Chinese varieties and the Chinese varieties were sub clustered into three subgroups (Xie et al., 2012). A similar study was also carried out on a total of 737 improved *indica* varieties collected worldwide. These collections were divided into two major groups with six subgroups using 384 SNPs and model-based population structure analysis (Wang et al., 2014). These results demonstrated the existence of population structure in *indica* rice, caused by selection for local ecological environments and spatial isolation. It is a common feature that varieties and advanced breeding lines of self-pollinated species have population

structure and familial relationship (Cockram et al., 2010a; Edae et al., 2014; Flint-Garcia et al., 2005; Pasam et al., 2012).

2.5.2.2 LD pattern of rice

The mapping resolution of association mapping depends on the LD structure of the association panel (population) used. The LD decayed to its half maximum within 65 kb for a population of 13 *indica* rice accessions genotyped with genome-wide SNPs (Xu et al., 2012). The squared correlation coefficient (r^2) dropped from 0.52 to 0.2 and 0.1 at 101 kb and 343 kb in a population of 114 Vietnamese *indica* rice accessions (Phung et al., 2014). The overall LD decayed to its half value (r^2~0.25) at around 300 kb in a collection of 220 rice accessions collected from various national and international institutes (Kumar et al., 2015). The genomes of 40 cultivated accessions selected from the major groups of rice and 10 accessions of their wild progenitors (*O. rufipogon* and *O. nivara*) were re-sequenced. LD decayed to its half maximum within less than 10 kb for *O. rufipogon* and *O. nivara*, 65 kb for *indica* and 200 kb for *japonica*. For subpopulations within *japonica*, LD was also high, with the half-maximum at ~300 kb, 300 kb and 180 kb for *aromatic*, *temperate japonica* and *tropical japonica*, respectively (Xu et al., 2012). The difference in the extent of LD observed in different genomic regions and different population could probably due to the difference in recombination rate, selection history or structural difference among populations. There is usually large variance associated with an estimate of LD decay, which is partially caused by uneven marker distribution and the difference in marker minor allele frequency (MAF).

2.5.2.3 Association studies using sparse markers

The first association study in rice was conducted by Agrama et al. (2007) using 92 rice germplasm from seven geographic regions of Africa, Asia, and Latin America, and eleven US cultivars genotyped with 123 SSR markers. Many of the associated markers were located in regions where QTL had previously been identified. De Oliveira Borba et al. (2010) applied association analysis to identify eight marker-trait associations (MTAs) for four yield and grain quality traits using the Embrapa Rice Core Collection genotyped with 86 SSR markers. The marker RM190 is associated with amylose content (AC) across years and cultivation. Jia et al. (2012)used the 217 sub-core entries from the USDA rice core collection genotyped with 155 genome-wide markers to identify markers associated with sheath blight (ShB) resistance. Ten marker loci on seven chromosomes were identified to significantly associate with a

response to the ShB pathogen. The same rice accession panel was also used to map QTL for GY and other traits (Li et al., 2011). Thirty MTAs were highly significant, including four for GY. Allelic analysis of OSR13, RM471 and RM7003 for their co-associations with yield traits indicated that allele 126 bp of RM471 and 108 bp of RM7003 had the greatest positive effect on yield traits.

Courtois et al. (2011) conducted AM using a set of 305 varieties from the European Rice Germplasm Collection with 90 SNPs. No MTAs were found for duration and grain type due to the overlap between the genetic and phenotypic structure. Associations were found for other traits including salinity tolerance. Using 416 rice entries including landraces, cultivars and breeding lines collected mostly in China and genotyped with 100 SSR markers, Jin et al. (2010) found that the Wx and starch synthase IIa (SSIIa) genes were strongly associated with apparent amylose content (AAC) and pasting temperature (PT). They also found that five and seven SSRs were strongly associated with AAC and PT, respectively.

2.5.2.4 Candidate gene-based association mapping

A candidate gene-based association mapping within a European Rice Core collection (ERCC) comprising 180 *japonica* elite lines was carried out using 124 SNPs for 47 candidate genes and 52 SSRs for 22 candidate genes and 14 QTLs for salinity tolerance in rice (Ahmadi et al., 2011). A total of 19 distinct loci significantly associated with one or more salinity response traits were detected. Cloned HD controlling genes, such as *Hd1*, *Ghd7* contain a CCT domain. Only six of the 41 CCT family genes have been confirmed to control HD in rice. Using candidate gene-based association mapping method, 19 out of the 41 CCT domain-containing genes were identified to regulate HD in a collection of 529 rice accessions. Two of the associated genes were confirmed by transformation method (Zhang et al., 2015).

2.5.2.5 GWAS

The development of high density SNP chips in rice made real GWAS possible. An Affymerix SNP array containing 44,100 SNPs (44k chip) was used to genotype a collection of 413 diverse rice accessions collected from 82 countries, which was also systematically phenotyped for 34 traits (Zhao et al., 2011). GWAS using MLM identified dozens of common variants influencing numerous complex traits, including the previously known ones. 383 diverse rice accessions genotyped with the 44k chip was used to identify MTAs for Al tolerance based on root growth (Famoso et al., 2011). A total of 48 genomic regions

associated with Aluminum (Al) tolerance were identified. Most of the associations were subpopulation-specific. Four of these regions co-localized with a priori candidate genes, and two genomic regions co-localized with previously identified QTL. Bi-parental QTL mapping results showed three genomic regions corresponding to induced Al-sensitive rice mutants (*ART1*, *STAR2*, *Nrat1*). Susceptible and tolerant haplotypes were identified around *Nrat1* gene, which explained 40% of the Al tolerance variation within the *aus* subpopulation. Sequence analysis of *Nrat1* identified a trio of non-synonymous mutations predictive of Al sensitivity in the panel. GWAS discovered more MTAs with higher resolution, but bi-parental QTL mapping identified critical rare and/or subpopulation specific alleles that GWAS failed to detect. The primary and total root growth phenotypic data from the 233 rice accessions genotyped with the 44k Chip were used by (Zhao et al., 2011) to identify MTAs. Two genomic regions for primary root growth and four regions for total root growth were identified in the whole population. A collection of 315 accessions genotyped with the 44K chip and phenotyped in Yangzhou of China and Arkansas of America (Zhang et al., 2015) identified seven, five, 10, eight and six genomic regions significantly associated with panicle length, PB, GL, GW and grain length/width ratio, respectively.

Huang et al. (2010) identified around 3.6 million SNPs and construct a high-density haplotype map of the rice genome by sequencing 517 rice landraces and using a novel data-imputation method. The GWAS performed for 14 agronomic traits using the compressed MLM identified 37 significant associations for the tested 14 traits, of which six QTL were located close to previously known genes. This approach was extended to a larger and more diverse panel of 950 worldwide rice varieties, including the *indica* and *japonica* subspecies (Huang et al., 2012). A total of 32 new MTAs with FT and 10 grain-related traits were identified. A subset of 366 *indica* from the 517 rice accessions sequenced by Huang et al. (2010) was used to identify 30 markers associated with blast resistance (Wang et al., 2014).

2.6 Breaking the yield barrier through breeding

Rice breeders today face the challenge of how to increase productivity by effectively integrating well-established conventional breeding methods with new approaches offered by rapid advances in molecular marker technology and genomics. Breaking the yield barrier can be achieved in two ways: Increasing “attainable yield” such as from insect resistance, etc., and from “potential yield” such as raising the base yield due to changes in physiological processes. Increasing attainable yield can be achieved by mining super alleles for

resistance/tolerance to biotic and abiotic stresses using genomics-enabled methods and introgressing and/or pyramiding the identified alleles in elite lines through MAS. Increasing yield potential is more challenging, since more genes with minor effects must be used simultaneously. An integrated breeding strategy has been proposed for the irrigated rice breeding program at IRRI to break yield barrier (GRiSP, 2010). The strategy utilizes recurrent selection (RS) to quickly pyramid the major QTL that have been proven useful in the breeding population in the first few selection cycles and maintain the genetic variation contributed by the many minor genes to be explored in later cycles and explores GS for reducing the length of breeding cycles and the cost of expensive phenotyping. It is necessary for the success of the integrated strategy to strengthen the research and development in following areas. Considering that the public acceptance of transgenic product is low in most of rice production countries, genetic engineering is not seen as a viable option in the near feature.

2.6.1 Increasing genetic diversity of breeding gene pools

The success of a breeding program depends on having adequate diversity in the germplasm. Domestication has reduced genetic diversity in rice relative to their wild progenitors. Modern cultivated rice is estimated to retain only approximately 10-20% of the genetic diversity present in its wild rice ancestor, *O. rufipogon* (Caicedo et al., 2007; Zhu et al., 2007). As advanced breeding stocks and materials are generated, one casualty is the diversity itself. There is growing concern that limited genetic diversity in present day high-yielding rice germplasm impedes further improvement in productivity. It is feared that unless new diversity is infused into the breeding germplasm, we face catastrophic reductions in crop productivity if the climate turns adverse.

There are several options of increasing genetic diversity of a breeding program. The first option is the introduction of elite cultivars and breeding lines from other programs. IRRI's materials have been introduced into many national breeding programs, including Philippines, India, Bangladesh, and China, etc. There was also introducing of materials between national breeding programs. An extreme form is to combining diverse alleles from *indica* and *japonica*. Yonemaru et al. (2014) reported that the combination of *indica* and *japonica* factors appears to have the greater potential for increasing yield because the admixture-type cultivars were most prevalent in high-yielding Japanese cultivars. The majority of Chinese 'super' rice varieties are admixture-type (Chao et al., 2010).

The second option is to use genetic variability exists in cultivated rice, although this is just 5% of the total variability existing in the genus *Oryza* (Tanksley and McCouch, 1997) . Landraces contain large amount of useful genes conferring resistance to abiotic and biotic stresses (Jackson, 1997; Pusadee et al., 2009; Shi et al., 2010; Thomson et al., 2007). Landraces have contributed many elite genes or alleles to the modern rice breeding programs, including the semi-dwarf genes (Monna et al., 2002; Sasaki et al., 2002), resistant/tolerant genes to diseases and pests, and genes for improving GY and quality (Fukuoka et al., 2009; Ishimaru et al., 2013). Therefore, it is likely that landraces may be important genetic resources for future rice genetics and breeding. An extreme form of combining diverse alleles from *indica* and *japonica* is one way to produce desirable genotypes for high-yielding rice. Yonemaru et al. (2014) reported that the combination of *indica* and *japonica* factors appears to have the greater potential for increasing yield because the admixture-type cultivars were most prevalent among high-yielding Japanese cultivars.

The third option of increasing genetic diversity of a breeding program is to use wild and exotic relatives of rice. Wild relatives hold a wealth of alleles that, if we can identify them, can help break yield barriers and enhance tolerance to various abiotic and biotic stresses. Plant breeders have recognized the worth of wild species and used these for improving simply inherited traits. The most successful examples of utilizing wild species include the use of *O. spontanea* as a source of wild abortive cytoplasmic male sterility (Li and Zhu, 1988), *O. nivara*genes to provide resistance against grassy stunt virus (Khush et al., 1977a), and *O. longistaminata* gene *Xa21* (Ikeda et al., 1990; Khush et al., 1990), *O. rufipogon* gene *Xa23* (Zhang et al., 2001) and *O. nivara* gene *Xa38* (Bhasin et al., 2012) for resistance against bacterial blight. However wild and exotic relatives are generally difficult to use. There is difficulty in utilizing wild species due to the incompatibility between wild species and modern cultivars (Brar and Khush, 1997). Furthermore, a successful cross brings in all of the deleterious factors along with the genetic factor of interest. Therefore, crosses between adapted varieties and exotic accessions require a long time to derive useful genetic material. Breeders are reluctant to use exotic germplasm in their breeding programs.

The extensive availability of molecular markers allows the monitoring of segments of the genome to facilitate the introgression of chromosome blocks. The identification of introgressed markers followed by the subsequent comparison with the whole-genome sequence will identify the chromosomal regions that are now transferred from one strain to another. Following the availability of advanced-backcross quantitative trait loci (AB-QTL) approach proposed by Tanksley and Nelson (1996), discovery of favorable genes from wild

species received impetus as a plant breeding strategy. Incorporation of potentially favorable alleles from wild ancestors of rice into improved genotypes for productivity traits has emerged as a promising strategy. Consequently, several yield-enhancing alleles from wild species of rice, such as *O. rufipogon*, were identified and introduced into high-yielding elite cultivars (Jin et al., 2009; Luo et al., 2013; Moncada et al., 2001; Thomson et al., 2003; Xiao et al., 1996; Xiao et al., 1998; Xie et al., 2006). Apart from *O. rufipogon*, search for yield-enhancing QTL alleles was successfully extended to *O. glaberrima* (Aluko et al., 2004), *O. grandiglumis* (Yoon et al., 2006), *O. glumaepatula* (Rangel et al., 2008), *O. longistaminata* (Chen et al., 2009) and *O. minuta* (Gaikwad et al., 2014; Linh et al., 2008).

2.6.2 Mining desirable alleles

Among diverse germplasm, there may be several allelic variants based on differences in the nucleotide sequence that arose during the history of rice domestication and differentiation. Such variants may have functional genetic differences. These distinct alleles will remain hidden unless efforts are initiated to screen these alleles for their potential use and function. The process of identifying alleles of a known gene/locus that are involved in a particular mechanism for any given trait and their variants within other genotypes or identifying novel, superior and beneficial alleles from the germplasm or natural population is known as allele mining. With the availability of high quality sequenced genomes of rice, allele mining provides the avenue for the validation of specific gene(s) responsible for a particular trait and mining of the most favourable alleles from the rice genebank.

The allele mining approach, based on the sequencing of different alleles of a single gene in different genotypes has been applied to mine the sequence diversification at the level of the key genes. Fukuoka et al. (2009) identified 12 variants (haplotypes A to L) in a set of cultivars that represented the range of genetic variation within cultivated rice on the basis of insertion- deletion polymorphisms at three positions in a proline-rich region. A series of chromosomal segment substitute lines, each of which possessed one of the *pi21* haplotypes in the genetic background of a susceptible cultivar, found that only the line carrying haplotype L showed improved resistance to blast. The allele mining approach, based on the sequencing of different alleles of a single gene in different genotypes has been applied to mine the sequence diversification at the level of the key genes. Twenty-four SNPs and one InDel mutation obtained through the comparison of the Starch Synthase II (SSII) gene sequence in 30 rice genotypes (Bao et al., 2006). Moreover, Takanokai and co-workers (2009) compared the

sequence of the GS3 gene, responsible of grain size, in 54 rice cultivars and identified 86 SNPs and 20 InDels. The allele mining for the rice fragrant gene badh2 also allowed the development of diagnostic molecular markers (Shi et al., 2008). Allele mining experiments applied to the Waxy gene led to the identification of five different allelic variants (Mikami et al., 2008). The re-sequencing of the *Waxy* gene and 1 kb of the putative upstream regulatory region performed in 21 genotypes representing all the apparent amylose content (AAC) classes identified previously identified several previously un-characterized SNPs were identified and four of them were used to develop dCAPS markers (Biselli et al., 2014).

Negrão et al. (2013) identified 15, four, six, six, and nine haplotypes for five of the key genes related to salt tolerance including *OsCPK*17 (*Os07g06740*), *OsRMC* (*Os04g56430*), *OsNHX*1 (*Os07g47100*), *OsHKT1;5* (*Os01g20160*) *and Sal*T (*Os01g24710*), respectively, by genotyping 392 rice accessions with EcoTILLING. Association analyses indicated that 11 significant SNPs associated with salinity. Platten et al. (2013) identified seven major allele groups of , *OsHKT1;5* within *O. sativa*, and comparison of leaf Na^+ concentrations across a number of diverse landraces allows a tentative hypothesis to be proposed as to the relative strength of the various alleles: Aromatic > aus >Hasawi>Daw>Agmi>IR29>Japonica (Platten et al., 2013).

2.6.3 Adopting recurrent selection

Grain yield is controlled by many genes of small effects and affected by many environmental factors collectively. The favorable alleles are likely to be spread across more than two lines, therefore requiring the assembly of alleles from different sources in a single inbred line in order to achieve significant improvement. Recurrent selection (RS) is a cyclical breeding strategy involving three main steps: selection, evaluation and recombination of the best performing selections. Multiple genotypes are intermated to increase the chance of creating novel allelic combinations. Through multiple cycles of recombination between selected genotypes linkage blocks are broken down and favourable genes are accumulated (Hallauer, 1985).

RS is an important breeding method for open-pollinated species, such as maize. It has been successfully used in many self-pollinated cereal crops, including rice. Chinese breeders initiated systematic wheat RS programs in 1992. Male sterility lines, which were controlled by dominant nuclear gene, were used to facilitate intermating. Using RS method, more than 35 commercial wheat varieties have been released (Ye et al., 2013). The average yield of the

best variety increased by 19%. At the University of Queensland, the best advanced wheat lines extracted from the second cycle of a RS program produced 15%-20% more grain yield than the best local commercial varieties in MET (Mark Dieters, personal communication). RS has been adopted in barely breeding programs. More than 50 % of acreage in Saskatoon, Canada grows barley varieties developed by a RS program. RS has been successfully used in upland rice breeding programs in France, CIAT and a few other countries. The best advanced line from the CIAT program could produce 35% more yield than the best commercial check (Gurimaraes, 2005). Male sterile lines have been adopted in most RS programs in self-pollinated crops to simplify intermating process, though manual crossing has been used by few programs.

2.6.4 Targeting cultivars to production environment

Plants are remarkable because of their ability of adapting themselves to surrounding environments to survive. The phenotype of a rice cultivar varies greatly in different environments. This response can be viewed as phenotypic plasticity, which is the ability of a genotype can produce multiple phenotypes in response to the environment (Des Marais et al., 2013). A genotype will display fluctuated yield production under different environments, which are referred to as GEI (Allard and Bradshaw, 1964; Kang, 2004). Genotype, environment and GEI collectively determine the phenotypic performance of a variety (Falconer and Mackey, 1996). GEI reduces the genetic gain in plant breeding programs through minimizing the association between phenotypic trait and genotypic values (Comstock and Moll, 1963). GEI could be either exploited by selecting superior genotype for a specific target environment or avoided by selecting widely adapted and stable genotype across wide range of environments (Kaya et al., 2006; Mitrovic et al., 2012). A variety with stable yield performance across varying environmental conditions could be planted in large area (exploring the general adaption). On the other hand, to obtain the best productivity requires suitable varieties to their optimal growing environments (utilizing the specific adaptation). To improve crop performance it is necessary to define the target population of environments (TPE) and to adopt the most suitable breeding strategy. TPE is the set of environments, fields, and seasons in which improved varieties are expected to perform well and stably (Comstock, 1977). These environments vary in predictable ways such as annual rainfall patterns, toposequence, soil type, and cultural practices and in unpredictable ways such as random drought or disease incidence (Lafitte et al., 2003). Testing candidate genotypes in multiple

environments of the TPE is commonly practiced by breeders to evaluate the general and specific adaptation of the candidate varieties and the importance of GEI.

GEI for GY of rice grown in rainfed lowland ecosystems has been extensively studied in different countries (Cooper et al., 1999; Ouk et al., 2007; Tariku et al., 2013; Wade et al., 1999), especially in Asia, the major rice production area. Results showed that GEI was large and more important than the genotypic main effects (Cooper and Somrith, 1997; Cooper et al., 1999; Henderson et al., 1996; Inthapanya et al., 2000; Wade et al., 1997). In contrast, the information of GEI for GY of irrigated lowland rice is less. For the irrigated lowland rice, Samonte and Hernandez (1990 and 1991) found that season-by-location interaction (SLI) was highly significant in the combined analyses of variance over seasons and locations, while locations and seasons were not significant. Genotype-by-season interaction and genotype-by-location interaction were found significant only in 10 and seven out of the 48 combined analyses, respectively, while genotype by-season-location interaction was significant in almost all cases. Genotype-by-year interactions were mostly non-significant for both dry and wet seasons but genotype-by-year-by-location interactions were mostly significant for both seasons. It should be pointed out that relatively small number of highly selected common genotypes (27-30) was used for each of the combined analyses. The international network for genetic evaluation of rice (INGER) conducted GEI analysis using a few datasets from the international irrigated rice yield nursery (IIRYN) conducted in 1993, 1994 and 1995. Combined analysis of variance conducted for each of the nurseries (early maturity or late maturity nurseries) suggested that the GEI sum of squares was 3-7 times of the genotype sum of squares variance (INGER, 1993a, 1993b, 1994a, 1994b, 1995a, 1995b). However, the large GEI could be at least partially caused by the acidic soil or biotic or abiotic stress conditions in some of the testing sites. Therefore, there was only limited information on the magnitude of GEI for GY of rice applicable to a more diverse set of genotypes that has not undergone intense selection for yield stability across diverse environments.

2.6.5 Increasing phenotyping precision

The majority of information breeders use to make selection decision is collected from field trials. Furthermore, precision phenotyping is also critical to the success of QTL mapping or association mapping experiments. High throughput genotyping is of little value if the phenotype is not accurately measured. The phenotyping component is the bottleneck to further improvement as it is often difficult, expensive, time consuming, laborious and

technically difficult and it is often destructive. Therefore, it is important to obtain the maximum and most accurate information from phenotyping trials. An efficient experimental design results in higher precision of the estimates of the treatment (genotype) effects, reduced residual variance and increased heritability (Clarke and Stefanova, 2011; Federer and Wolfinger, 2003; Whitaker et al., 2001). Many statistical methods have been developed to account for within-block heterogeneity at the data processing stage and to exploit 'indirect' information provided by genetic relatedness between genotypes and correlation between testing environments (Muller et al., 2010; Piepho et al., 2008; Smith et al., 2005). It is necessary to combine efficient experimental design and sophisticated statistical analysis methods to remove/reduce within-block heterogeneity and increase precision of experiment and heritability.

Non-invasive technologies that measure plant form, function, aspects of metabolism and content, often called phenomics, has been touted as a way to overcome the phenotyping bottleneck as measurements are often fast, repeatable and accurate and can be automated (Furbank and Tester, 2011). Recently, technologies to facilitate the observation and monitoring of different environmental, biophysical and physiological conditions in the field have been emerging. Continuous acquisition of data on all major inputs and outputs from farm-scale experiments allows breeding lines being evaluated for their yield potential under changing environmental conditions. By correlating all data with phenotypic changes and particular responses of the genome and gene expression to a dynamic environment more predictive models can be developed for predicting the performance of genotypes to environments (Fiorani and Schurr, 2013).

2.7 Conclusion

To improve the productivity and stability of rice, breeding for high yielding and durable abiotic and biotic stress-tolerant cultivars is essential. To enable MAS and understand the molecular mechanisms of important agronomic traits, great efforts have been devoted to the dissection of the genetic basis of yield and related traits. Traditional QTL mapping using biparental populations has identified numerous QTL for yield related traits, few of which have been well characterized. Only a small portion of those detected QTL has been applied in practical breeding programs. The reasons to explain this were discussed by Guo and Ye (2014). One reason is the persistence of the linkage phase between the target QTL and its linked markers across multiple populations. Markers linked to the QTL identified through

linkage mapping using one or a few populations may not be useful in gene pyramiding because different subsets of QTL will be polymorphic in each population, and the linkage phases between a marker and QTL alleles can differ even between closely related genotypes. Another reason is that the targeted QTL are mapped only coarsely by using small conventional bi-parental mapping populations and a limited number of markers. The transferred large QTL region might hold several genes. The recombination between those genes will then modify the effect of the targeted QTL. Furthermore, unfavourable linkage drag may be caused by the unintentional introduction of undesirable alleles.

Advance in NGS technologies reduces the genotyping cost and allows the use of abundant markers and large populations to fine-map the QTL of interest and develop functional markers subsequently. An international effort has been made to complete sequencing 3,000 rice genomes with an average sequencing depth of 14× (The 3000 Rice Genomes Project, 2014). Successful GWAS for yield related traits using different populations consisting of varied number of rice accessions genotyped with high density markers has been reported (Huang et al., 2010, 2012; Zhao et al., 2011). The resulting information will benefit to understand the genomic diversity within rice at a higher level of detail and will facilitate breeders to breed new rice varieties with high yielding, stability and multiple disease resistance.

The overall objective of the research reported in this thesis is to provide essential phenotypic and genetic information directly relevant to future breeding for irrigated ecosystems. A population of 392 cultivars or advanced breeding lines mostly from the IRRI, PhilRice (Philippines) and a few breeding programs of other countries collected to be used as the one of the base populations of IRRI's irrigated breeding program was used to evaluate GY and related traits in eight environments including Jiangxi (JX) and Sichuan (SC) in China, and six season (2) and nitrogen rate (3) combinations at IRRI headquarters (Los Baños, Philippines). 360 lines were used to identify patterns of genotype, environment, and GEI for GY (Chapter 3). This population was genotyped with 46 markers of 39 well characterized genes/QTLs for yield and related traits to test their usefulness and effectiveness through association analysis and investigate prediction of GY with markers for these known genes/QTLs using multiple regression analysis (Chapter 4). The population was genotyped with GBS, GWAS for yield and related traits was carried out to identify new MTAs (Chapter 5). The results from these researches would provide breeders with useful information and assist them to design more efficient breeding strategies.

Chapter 3 Genotype-by-environment interaction is important for grain yield in irrigated lowland rice

Abstract

Irrigated rice contributes the most to rice production globally and in most of rice producing countries. As with other rice production ecosystems, genotype and growing environment are the main factors influencing grain yield (GY) in irrigated rice production. However, the relative importance of genotype, environment and genotype-by-environment interaction (GEI) on GY is less studied for irrigated rice, although evaluation of candidate varieties for yield stability through multi-environment trials before release and registration is mandatory in many countries. To identify patterns of genotype, environment and GEI for GY of irrigated lowland rice, 392 cultivars or advanced breeding lines mostly from the International Rice Research Institute (IRRI), PhilRice (Philippines) and a few breeding programs of other countries were evaluated for GY and related traits in eight environments including Jiangxi (JX) and Sichuan (SC) in China and six season (2) and nitrogen rate (3) combinations at IRRI headquarters (Los Baños, Philippines). A wide range of variations across genotypes and environments were observed for all traits. Genotype, environment and GEI all significantly affected GY and some of the yield associated traits. GEI was more important than genotypic main effect for GY, seed setting rate and the number of panicles per plant but less important for other traits. For GY, the genotype-by-season interaction and genotype-by-season-by-nitrogen interaction was more important than the genotype-by-nitrogen interaction. The 303 genotypes with GY data in all environments were clustered into 10 groups based on GY using an agglomerative hierarchical clustering procedure. The eight environments were grouped into three groups based GY using the additive main effects and multiplicative interaction (AMMI) analysis. Three nitrogen rates in the WS and SC were grouped together (E1), while three nitrogen rates in the DS formed another group (E2). JX alone was the third group (E3). Genotype groups (GG) GG9 and GG3 had the highest GY across the eight testing environments. GG9 had the highest yield in E1 and E2 while GG10 was the best in E3.

Keywords: Additive main effects and multiplicative interaction (AMMI) analysis, Grain yield, Genotype-by-environment interaction (GEI), rice (*Oryza sativa* L.)

This chapter has been published in the journal of Field Crops Research. Abstract is formatted according to the journal standards

3.1 Introduction

Rice (*Oryza sativa* L.) is one of the major food crops in the world and generally extensively consumed in developing countries. More than half of the people on the globe depend on rice for their basic diet. The world's population is expected to increase by about two billion in the next two decades and half of the increase will be in Asia (Gregory et al., 2000). It is estimated that the global rice production must reach 800 million tons in 2025 (currently world paddy production in 2014 is 744.4 million tons) (FAOSTAT, 2014) to meet the demand for rice consumption.

Rice is grown under a wide range of environmental and climatic conditions ranging from lowland to upland and irrigated to rainfed situations. Further increase in rice yield production requires the development of high yielding genotypes with desirable agronomic traits and stability for diverse ecosystems (IRRI, 2006). The irrigated rice area accounts for about 56 percent of the total area and contributes 76 percent of the total production in the world (Papademetriou et al., 2000), which was the source of the large increases of productivity leading to the Green Revolution (Fischer et al., 2012). It is important for food security in many countries including China to increase the productivity of irrigated rice by developing new varieties and adopting improved agronomic practices.

Yield performance of different rice genotypes vary greatly in accordance with the environment. Genotype, environment and the genotype-by-environment interaction (GEI) collectively determine the phenotypic performance of a variety (Falconer and Mackey, 1996). The stability of yield performance is one of the most desirable characters of a genotype to be released as a variety, which allows the developed varieties to be adopted in large area (exploring the general adaption). On the other hand, to achieve maximum productivity requires targeting varieties to their best growing environments (utilizing the specific adaptation). Testing candidate genotypes in multiple environments of the target population of environments (TPE), which is a set of environments in which improved varieties developed by a breeding program need to be adapted (Comstock, 1977), is commonly practiced by breeders to evaluate the general and specific adaptation of the candidate varieties and the importance of GEI.

GEI for GY of rice grown in rainfed lowland ecosystems have been extensively studied in different countries (Cooper et al., 1999; Ouk et al., 2007; Tariku et al., 2013; Wade et al., 1999), especially in Asia, the major rice production area. Results showed that GEI was large and more important than the genotypic main effects (Cooper and Somrith, 1997; Cooper

et al., 1999; Henderson et al., 1996; Inthapanya et al., 2000; Wade et al., 1997, 1999). In contrast, GEI for GY of rice grown in irrigated lowland has not received adequate attention comparable to its importance. The possible reasons might be (1) Rice production under well managed irrigated environments does not usually have the problem associated with water availability, which varies greatly in dryland crop production; (2) The huge success of IRRI's irrigated breeding program in developing varieties for many countries by simply conducting all the major selections at its headquarters may have led to the ignorance of possible GEI for GY under the favorable irrigated ecosystem. Nevertheless, stable performance in a wide range of environments is one of the important criteria for new variety release in many countries. Samonte and Hernandez (1990 and 1991) analyzed data from the Irrigated-Lowland Rice National Cooperative Testing (NCT) Program of the Philippines, which included four maturity subgroups tested in 8 seasons at 10 locations. They found that season-by-location interaction (SLI) was highly significant in the combined analyses of variance over seasons and locations, while locations and seasons were not significant. Genotype-by-season interaction (GSI) and genotype-by-location interaction (GLI) were found significant only in 10 and seven out of the 48 combined analyses, respectively, while genotype by-season-location interaction (GSLI) was significant in almost all cases. Genotype-by-year interactions (GYI) were mostly non-significant for both dry and wet seasons but genotype-by-year-by-location interactions (GYLI) were mostly significant for both seasons. It should be pointed out that relatively small number of highly selected common genotypes (27-30) was used for each of the combined analyses. The international network for genetic evaluation of rice (INGER) coordinated by IRRI conducted GEI analysis using a few datasets from the international irrigated rice yield nursery (IIRYN) conducted in 1993, 1994 and 1995. These nurseries had many testing sites (20-45) in many countries (9-22) while the numbers of genotypes included were relatively small and from many breeding programs (27-28). Combined analysis of variance conducted for each of the nurseries suggested that the GEI sum of squares was 3-7 times of the genotype sum of squares variance (INGER, 1993a, 1993b, 1994a, 1994b, 1995a, 1995b). However, the large GEI could be at least partially caused by the fact that the soil and weather conditions in some of the testing sites were untypical. For the 1993 early maturity nursery, the crop at Gazipur (Bangladesh) was submerged under stagnant water for nine days, the trial at Alor Setar (Malaysia) was grown in acidic soil (PH4.5) and at the Malan (India) site experienced low-temperature stress at flowering stage (INGER, 1993a). For the 1993 medium maturity nursery, the Gazipur (Bangladesh) and Tando Jam (Pakistan) sites were planted under rainfed conditions, the Alor

Setar and Seberang Perai (Malaysia), the Binh Quoi (Vietnam) and Batalgoda (Sri Lanka) sites were in acidic soils (PH ranging 4.5 to 5.7). Aduthurai (India) site was submerged for more than one week due to heavy rains (INGER, 1993b). For the 1994 early maturity nursery, the Ngale (Indonesia), Cantho (Vietnam), Gazipur (Bangladesh) and Port Blair and Karjat (India) were planted under rainfed conditions (INGER, 1994a). Therefore, there was only limited information on the magnitude of GEI for GY of rice applicable to a more diverse set of genotypes that has not undergone intense selection for yield stability across diverse environments.

Nitrogen is one of the most yield limiting nutrients in rice production. Farmers often apply a higher amount of nitrogen fertilizer than the minimum required for maximum crop growth (Lemaire and Gastal, 1997). However nitrogen application cannot promise a substantial increase in crop productivity due to the principle of diminishing returns (Cassman et al., 2003) and can cause serious nutrient pollution (Zhao et al., 2012). Therefore it is important to understand the interactions between the nitrogen application and rice yield. The irrigated rice breeding program at IRRI has a mission to develop new rice germplasm that is widely adapted and produces high and stable yields across a broad range of environments in less developed countries. IRRI germplasm has been directly released for production or used as crossing parents in many breeding programs in different countries, including China. The INGER irrigated nurseries sometimes include one or two testing locations in China. However, in the last about 15 years, IRRI's breeding lines were not systematically tested in China and exploited by Chinese breeders, partially because cultivars developed using Chinese breeding materials are more adaptable to the subtropical production environments in China and IRRI's irrigated breeding has been targeting tropical regions. However, Chinese rice breeders recently showed renewed interested in ultilizing IRRI's breeding lines to increase the genetic diversities of their breeding programs. IRRI breeders are interested in ultilizing Chinese breeding lines to futher increase yield potential. A joint effort to exploit the complemetary characteritics of the Chinese and IRRI breeding gene pools for developing better germplasms for South and Sout-east Asia is under discussion. Knowledge of the effect of GEI on GY is required for the design of an efficient and economic selection strategy involving a shuttle component (Gauch and Zobel, 1996; Kang, 1998). Thus it is essential to understand the responses of IRRI cultivated rice varieties in targeted Chinese locations and the performance of Chinese rice lines at IRRI.

By testing a large set of advanced rice breeding lines and cultivars in two target locations in China and two distinct seasons under three different rates of nitrogen application

at IRRI, the present study investigated the effects of genotype, environment including location, season and nitrogen rate and GEI on GY and related traits under irrigated conditions. The two Chinese locations, Jiangxi (JX) and Sichuan (SC), are large *indica* rice production areas with distinct soil and climatic conditions. Our experiences suggested that genotypes performed well in JX and/or SC tend to have good adpation in other *indica* rice production areas in China as well. The results of the present study were used to explore opportunities for selecting breeding lines in IRRI for the two targeted locations in China.

3.2 Materials and methods

Plant materials

Three hundreds and ninety two rice lines developed for the irrigated lowland ecosystem were used in this study to have a large genetic diversity. About 16% are released cultivars, while the rest are advanced breeding lines. Majority of the lines were from IRRI (223). The number of lines from PhilRice, CIAT, China and Vietnam were more than ten. The rest of the lines were from programs in Bangladesh, Colombia, Indonesia, Nepal, Africa Rice Center, Egypt and Pakinstan, India, Italy, Repubilic of Korea, Sri Lanka, Suriname, Taiwan and Turkey. The seeds were obtained from the International Network for Genetic Evaluation of Rice (INGER) and IRRI breeders (Appendix Table S1.). Since the present study is part of the population improvement project initiated in IRRI to broaden the genetic dveristy of IRRI's irrigated rice breeding populations, some of the lines did not perform well in the tropical environments (Philippines) and subtropical environments (China) and were removed from the GEI analysis (see the data analysis section below).

Testing environments

Jiangxi (JX) and Sichuan (SC) of China and IRRI headquarters (Los Baños, Philippines) were the three testing locations. JX and SC are two major rice production provinces in China with distinct soil and climatic characteristics. Genotypes perform well in JX and/or SC tend to have good adpation in other *indica* rice production areas in China as well. IRRI headquarters has been the major breeding site of IRRI's irrigated breeding program for more than 40 years (Table 3.1). The experiments in JX and SC were for one crop season in 2012. At IRRI, the experiment was carried in the dry season (DS) and wet season (WS) of 2012. Three nitrogen fertilizer application rates, no nitrogen, low (90 kg ha^{-1}) and high (180 kg ha^{-1}), were used to create three artificial environments in the DS, designated as DS1, DS2 and DS3, respectively.

Similarly, three nitrogen ferilizer application rates, no nitrogen, low (45 kg ha^{-1}) and high (90 kg ha^{-1}), were used to create three artificial environments in the WS, designated as WS1, WS2 and WS3, respectively. The use of no nitrogen application and releatively high nitrogen application (180 kg ha^{-1}) was to allow testing the nitrogen responses beyond the nomral nitrogen rate used in Philippines and other countries. Farmers in China ususally apply more than 180 kg ha^{-1}nitrogen while famers in many South-east Asia countries hardly apply any nitrogen. It has been suggested that IRRI should increase nitrogen rate in its DS trials to allow identifying breeding lines with higher yield potential (Dr. Akim Doberman, Personal communication). The nitrogen ferilizer application rate in the two Chinse locations was 150 kg ha^{-1}.

Trial description

A row-column design (28 × 14) with two replications was used for all the six environments at IRRI with different randomizations. Each plot was 2.56 m^2 consisting of 64 plants (8 × 8) with 20 cm spacing between rows and no vacant rows between plots. In the DS of 2012, the genotypes were seeded in a seedling bed on 22 and 23 November, 2011 and transplanted with a single plant per hill on 15 December, 2011 in the field of IRRI campus (Table 3.1). Two days before transplanting, Furadan was applied to control the golden snail and basal nitrogen was applied accordingly. Three weeks later after transplanting, the plants infected by tungro disease were removed by hand from the field to prevent disease spread. Nitrogen fertilizer was top dressed at 14 and 40 days after transplanting. Hand weeding and pesticide application was done as needed. Bird scaring practices were applied from anthesis to harvest to prevent grain losses. Rodents were controlled by setting traps. In the WS, entries were seeded on 12 June and transplanted on 6 July 2012 (Table 3.1). The management practices followed were the same as in the DS. Harvest was done in accordance with the maturity of each variety and the earliest batch started from 24 September 2012 and lasted until 26 October 2012.

In China, the experiment design used was a randomized complete block design with two replications. In JX, the plot was 2.56 m^2 consisting of 64 plants (8 × 8) spaced at 0.2 m × 0.2 m, while the plots in SC was 0.96 m^2 consisting of 24 plants (4 × 6) plants spaced at 0.2 m × 0.2 m. Plots were managed conventionally following the established local normal practices.

The plot size used in the present study is much smaller than that recommended for yield testing by IRRI (IRRI, 2013). Adopting the recommended plot size is impractical for

trials with a large number of lines, because the cost and the chance to introduce extraneous errors are high. Whether the GY measured using a small plot had a strong correlation with that measured in large plot has never been systematically studied. However, when the objective is not to select the very best genotypes, which is the case of the present study, multi-row small plot is likely to be appropriate. Indeed, small plot size was adopted by most (if not all) of the reported studies involving more than a hundred genotypes. Small multi-row plot is also widely used in yield testing by Chinese rice breeders, although the last stage of the Chinese National Variety Test uses large plot.

Measurement of traits

Grain yield and related traits were evaluated following IRRI's standard evaluation system for rice (IRRI, 1996). Flowering date was the date when more than half panicles of each plot were flowering. Days to flowering (DTF) was the sum of days from seeding to flowering date. Plant height (PH) was measured as the average height of five plants (three plants from second row and two plants from third row) in cm from the soil surface to the tip of the tallest panicle (awns excluded). Tiller number (TN) was counted as the average of the five plants which were measured for the PH. The five plants harvested at maturity separately from the middle of each plot were used for measuring the following agronomic traits. (1) Panicle number per plant (PN): the average number of panicles per plant; (2) Spikelet number per panicle (SN): the average number of the filled grains and unfilled grains, measured using three panicles per plant; (3) Grain number per panicle (GN): the average of filled grain number, measured using three panicles per plant; (4) Seed setting rate (SR): the ratio of GN to SN; (5) Number of primary branches per panicle (PB) was the average number of primary branches of three panicles from each plant; (6) Number of secondary branches per panicle (SB) was the average number of secondary branches of 3 panicles from each plant. Thousand grain weight (TGW): average weight of 1, 000 filled grain, measured in grams, average over two samples of 100 grains taken from the bulk harvested grains from each plot. All the materials were threshed and dried in the oven for two days at 55 °C and then stored in cool room for 2 months. This allowed the moisture content of all samples to uniformly reach around 14%. Grain yield per plot (GY) was the sum grain weight of the bulk harvested plants and the five plants harvested sperately for data collection (totally 36 plants).

All traits were measured for the six experiments at IRRI. GY, DTF, GN, PH, PN, TGW and SR were measured in SC, while GY, DTF, GN, PH and SR were measured in JX.

Data analysis

Due to photoperiod sensitivity, insect or rat damage, some of the lines couldn't give any production in some environments. Although we aimed at *indica* breeding lines, 24 lines were later confirmed to be *japonica* and removed from the data analysis. A few lines were found to be outliers using a model-based population structure analysis implemented in STRUCTURE (Pritchard et al., 2000b) and multi-dimensional scaling and cluster analysis implemented in the R packages AWclust (Gao and Starmer, 2008) based on 50 SSR markers evenly distributed among all chromosomes and removed as well (results not shown). Finally, 303 lines had GY data in all testing environments were used in multi-site analysis. The number of lines used for analysis for different traits varied slightly.

All trials were separately analyzed by fitting an appropriate spatial model with rows and columns using PBTools (bbi.irri.org) and R (R Core Team, 2015). The field plot row and column positions were used as fixed covariates to partially adjust for the possible local field trend. The best linear unbiased estimations (BLUE) from the best-fit model were used as raw data for all subsequent analyses. Classification of genotypes was performed using an agglomerative hierarchical clustering procedure with squared Euclidean distance as the dissimilarity measure (Williams, 1976) and Ward's method, which uses incremental sums of squares as the clustering strategy (Ward, 1963).

PBTools (bbi.irri.org) and R were also used to perform the two-stage combined analysis to estimate variance components of different sources. For GY, the GEI was also decomposed into genotype-by-season, genotype-by-nitrogen and genotype-by-season-by-nitrogen interaction. The AMMI model (Gauch, 1988) was used in analyzing the GEI for GY. The AMMI model is a combination of analysis of variance (ANOVA) and principal component analysis (PCA). ANOVA is used to analyze the main effects while PCA decomposes the interaction into PCA axes.

The analytical model can be written as

$$Y_{ge} = \mu + \alpha_g + \beta_e + \Sigma_n \lambda_n \gamma_{gn} \delta_{en} + \rho_{ge} + \varepsilon_{ger}$$

Where Y_{ge} is the trait of genotype g in environment e; μ is the grand mean, α_g is the genotypes deviation from grand mean and the environment deviation β_e, λ_n is the eigenvalue of PCA axis n; γ_{gn} and δ_{en} are the genotype and environment PCA scores for PCA axis n; ρ_{ge} is the residual of AMMI model and ε_{ger} is the random error.

Biplot was used to visualize the AMMI results. Genotype group performance plot was constructed by plotting mean grain yields for genotype groups against environment groups based on the untransformed mean yield.

3.3 Results

Overall performance

A wide range of variations was observed for GY and yield components among the genotypes and across the environments (Table 3.1 and Appendix Table S1). The ranges among environments for average GY, TGW, GN, DTF, PH, PN, SN and SR were 470.1-846.0g, 24.5-27.4g, 87-120, 86-105 days, 85.9-109.7 cm, 7-11, 110-134 and 58.7%-89.7%, respectively (Table 3.1). In the five environments including DS1, DS3, WS1, WS2 and WS3, where PB and SB were recorded, the ranges for the two traits were 10-11 and 21-27, respectively. TN, only measured in the three WS environments, ranged from 7 to 9. The ranges among genotypes for average GY, TGW, GN, DTF, PH, PN, TN, SN, SR, PB and SB were 286.7-995.7g, 18.2-36.4g, 67-160, 72-104 days, 75.2-132.3 cm, 5-19, 5-16, 78-202, 71.2-91.0%, 5-13 and 8-39, respectively.

Single-trial analysis

Single-site analysis showed that there were significant variations among genotypes for GY in all the eight environments (Table 3.2). The use of advanced design, row-column design, helped to reduce residuals. The residual variances were larger for the IRRI environments compared to the Chinese environments. Trials with nitrogen applications had relatively lower precision, indicating that it was difficult to uniformly apply nitrogen to all plots. Considering the large number of genotypes tested the trials were of high quality. The heritability of GY and related traits in each environment is displayed in Table 3.1. The heritability of GY ranged between 0.59 and 0.95 among the eight testing environments. The heritability of yield related traits varied from 0 (to 0.99 (GN and SR in SC).

Table 3.1 Characteristics of the eight environments and the basic statistics (minimum-mean-maximum-sd-h^2) of grain yield and related traits in each environment

Env.[a]	DS1	DS2	DS3	WS1	WS2	WS3	SC	JX
Max.T[b]	38	38	38	35	35	35	35	37
Min.T[b]	27	27	27	26	26	26	14	12
Seed[c]	22/11/2011	22/11/2011	22/11/2011	12/06/2012	12/06/2012	12/06/2012	25/04/2012	01/06/2012
Tplant[d]	15/12/2011	15/12/2011	15/12/2011	06/07/2012	06/07/2012	06/07/2012	25/05/2012	20/07/2012
Lati.(N)[e]	14.2°	14.2°	14.2°	14.2°	14.2°	14.2°	30.7°	27.6°
Longi.(E)[e]	121.2°	121.2°	121.2°	121.2°	121.2°	121.2°	104.1°	113.9°
Elev.(m)[e]	27	27	36	27	27	27	500	203.6
N (Kg)[f]	0	90	180	0	45	90	150	150
GY (g)[g]	96.8-470.1-1237.4-163.6-0.76	141.7-778.1-1848.6-297.7-0.76	172.8-751.9-1369.7-196.9-0.60	210.8-530.1-743.0-75.2-0.67	401.6-697.1-920-90.7-0.59	401.3-695.1-991.1-100-0.64	499.8-846.0-1350.3-132.2-0.95	200.0-705.9-1372.1-238.5-0.58
DTF (D)	59-87-104-6-0.94	60-86-104-6-0.92	60-91-108-6-0.88	74-89-104-5-0.94	70-89-104-5-0.86	68-90-109-6-0.93	79-105-137-11-0.99	62-100-111-7-0.74
GN	50-97-163-17-0.62	NA[h]	25-99-188-22-0.68	61-107-185-20-0.76	61-112-186-19-0.63	68-120-193-21-0.69	51-102-203-24	46-87-178-21-0.72
PB	5-10-18-1-0.6	NA	4-10-15-1-0.65	7-10-14-1-0.66	7-10-17-1-0.60	6-11-18-1-0.71	NA	NA
PH (cm)	61.8-85.9-113.8-6.7-0.87	62.5-96.6-137.9-8.1-0.89	56.4-94.5-134.2-8.4-0.89	62.4-96.4-118.1-8.1-0.88	78.3-108.4-136.2-9.3-0.85	79.6-109.7-141.5-10.3-0.91	76.5-107.4-147.1-11.3-0.80	56.5-107.7-146.3-10.3-0.87
PN	4-7-25-2-0.59	5-10-39-4-0.76	4-11-30-3-0.67	4-7-12-1-0.65	4-9-19-2-0.53	5-9-24-2-0.52	5-10-15-2-0.64	NA
SB	6-27-54-6-0.55	NA	5-25-48-5-0.67	10-21-39-4-0.56	12-24-47-4-0.47	14-25-44-5-0.38	NA	NA
SN	60-110-179-20-0.65	NA	26-126-253-27-0.69	70-120-199-22-0.76	69-126-216-22-0.64	83-134-207-22-0.69	82-134-242-28-0.71	NA
SR	66.5-87.9-96.5-4.0-0.26	NA	57.1-79.5-94.3-6.1-0.56	73.0-89.7-97.6-3-0.00	75.9-88.6-95.2-3.5-0.2	78.1-89.1-94.4-2.6-0.35	47.5-76.3-92.3-8.8-0.99	10.6-58.7-86.1-15.9-0.70
TGW (g)	17.6-27.3-37.6-2.4-0.95	17.4-27.3-37.3-2.5-0.96	18.3-27.4-37.8-2.5-0.94	17.8-27.0-36.8-2.5-0.94	19.3-27.4-37.3-2.4-0.89	17.3-27.1-37.3-2.5-0.92	17.2-24.5-35.7-2.4-0.96	19.6-24.9-31.9-2.3-0.85
TN	NA	NA	NA	4-7-12-1-0.71	4-9-19-2-0.53	6-9-16-2-0.60	NA	NA

[a] Environments, SC, Chengdu, Sichuan, China; JX, Pingxiang, Jiangxi, China; WS and DS, the 2012 dry and wet seasons of IRRI headquarters (Los Baños, Philippines) with three nitrogen fertilizer application rates; [b] Max.T and Min. T, the maximum and minimum temperature in °C from seeding to harvest, respectively.

[c] Seed, seeding date; [d] Tplant, transplanting date; [e] Lati, Longi and Elev, latitude, longitude and elevation.; [f] N, nitrogen fertilizer application rate (kg ha^{-1}).

[g] GY, grain yield per plot (gram); DTF, days to flowering; GN, grain number per panicle; PB, number of primary branches per panicle; PH, plant height (cm); PN, number of panicles per plant; SB, number of secondary branches per panicle; SN, spikelet number per panicle; SR, seed setting rate; TGW, thousand grain weight (gram); TN, number of tillers per plant. The basic statistics, minimum-mean-maximum-sd-h^2 were given for each trait, h^2 was the broad-sense heritability; [h] NA, Not available

Table 3.2 Variance components for GY from single environment analysis

Source	Environment							
	DS1	DS2	DS3	WS1	WS2	WS3	SC	JX
Genotype	16760.9	51893.4	19180.6	3072.7	4316.7	5952.6	10170.1	53775.8
Rep	NA[c]	NA	NA	NA	NA	NA	540.5	141.8
Rep:Row[a]	542.8	923.1	1708.4	470.2	1069.9	1449.6	NA	NA
Rep:Col[b]	1411.5	3597.1	3577.8	889.0	3256.6	2502.8	NA	NA
Residual	10463.8	33427.8	26017.8	4076.2	6021.3	6677.2	13055.5	5719.8

[a] Design row within Rep

[b] Design column within Rep

[c] NA, Not available

Combined analysis

The combined analysis indicated that for GY the genotype and GEI account for 17.6% and 29.8% of the total variance, respectively (Table 3.3). The relatively high magnitude of the GEI variance relative to the genotypic component was not unexpected. The three locations spanned a broad range of environments including tropical and subtropical climates and the distinctive dry and wet seasons at IRRI and different rates of the most important yield-effecting nutrition, nitrogen. The interaction between genotype and season (GSI) accounted for 22.0% of the total variation, while the interaction between genotype and nitrogen (GNI) and interaction between genotype, season and nitrogen (GSNI) were 0.9% and 6.9% of the total variation. For PN and SR, the GEI variance was also much larger than the genotypic variance (Table 3.3). However, the residual for SR was very large, indicating phenotyping accuracy was too low to be reliable. For TGW, PH, PB, TN and SB, the genotypic variance was much larger than the GEI variance (Table 3.3), indicating that GEI was negligible. For DTF, GN and SN the GEI variance was similar to that of the genotypic variance (Table 3.3), indicating that GEI was less important but should be taken into consideration in breeding.

Table 3.3 Variance components from a combined analysis for GY and yield related traits

Source	GY	DTF	GN	PH	PN	SR	SN	TGW	PB	SB	TN
Genotype	5872.7	16.8	141.0	46.1	1.5	2.1	199.8	5.2	0.7	9.3	9.3
GEI[a]	9915.6	11.4	92.6	12.4	2.5	4.1	120.5	0.3	0.2	3.3	3.3
Residual	17536.8	6.1	261.8	21.6	4.0	20.8	328.9	0.9	0.9	23.9	23.9

[a] GEI, genotype-by- environment interaction

AMMI

AMMI analysis (Table 3.4) showed that the first principal component axis accounted for 49.6% of the total variation and the second accounted for 18.8%. The first two PCAs accounted for more than 68% of the GEI. Based on results from many studies in different crops, AMMI 1 model and the AMMI 2 model were used to explain the present data.

Table 3.4 Percentage of total variation of for GY accounted for by the principal components through AMMI analysis

Source	Df	SS	MS	Percent[a]	Acum[b]
PC1	308	25437623.0	82589.7	49.6	49.6
PC2	306	9642346.0	31510.9	18.8	68.4
PC3	304	6337453.0	20846.9	12.4	80.8
PC4	302	4110797.0	13611.9	8.0	88.8
PC5	300	3511453.0	11704.8	6.8	95.6
PC6	298	1309109.0	4393.0	2.6	98.2
PC7	296	942642.7	3184.6	1.8	100.0
PC8	294	0.0	0.0	0.0	100.0

[a] Percentage of sum of squares accounted for

[b] Accumulative percentage sum of squares accounted for

The AMMI 1 biplot (Figure 3.1a) indicated that most genotypes tended to have IPCA 1 scores of nearly zero and their mean GY were close to around 700g. The environments differed from each other not only for main effect, but also for their interaction effects except WS2 and WS3, which located in the middle area in Figure 3.1a.

The environments WS2 and WS3 had IPCA 1 scores near zero and hence had small interaction effects, indicating that all genotypes performed well in these two environments. Thus, these two environments were considered as the favourable environments for all the genotypes tested. Similarly, those genotypes with zero score on the first IPCA 1 were less influenced by the environments. Furthermore, those lines with above average yield and IPCA 1 score close to zero were considered as the stable and had general adaptation to all the environments.

Both genotype and environment with similar signs of IPCA 1 had positive interaction and thus higher yield of the genotype at that particular environment. For instance, JX and SC among the environments and entry 276 among the genotypes had negative IPCA 1 score and

above average yield, hence entry 276 was identified as specifically adapted to these two environments. Likewise, the genotypes 18, 58, 91, 114 and 161 had above average yield in DS2 and DS3 and positive IPCA 1 score and thus these environments were favorable environments for these six genotypes.

In the AMMI 2 biplot (Figure 3.1b), the environments WS1, WS2, WS3, DS1, DS3 and SC had short spokes and did not exert strong interactive forces, while JX and DS2 had long spokes and exerted relatively strong interactive forces.

The position and perpendicular projection of genotype points relative to environment vectors could be used to determine whether a cultivar was specifically adapted to a given environment. Genotypes that were positioned further along the positive direction of a vector tended to show higher grain yield, reflecting better adaptation to that environment (Kempton, 1984). The majority of the genotypes tended to gather around zero. Entry 91, 114 and 161 were positioned along the positive direction of DS2, suggesting that they specifically adapted to DS2. This result was consistent with the result of AMMI 1 biplot.

In the AMMI 2 biplot (Figure 3.1b) the maximum angle among the vectors of SC and the 3 IRRI WS environments was well below 90 degrees, indicating they ranked cultivars similarly and therefore formed one environment group, designated as E1. Likewise, the maximum angle among the vectors of the 3 IRRI DS environments was well below 90 degrees, indicating they discriminated the genotypes similarly and formed the second environment group, designated as E2. JX alone formed the third environment group. Comparing the environmental vector for JX and DS3, DS2 and SC revealed an angle of nearly 180 degrees. Genotype discrimination in JX and SC was expected to be in almost the opposite direction to that of DS3 and DS2 environment, respectively. Thus there appeared to be three distinct groups of environments discriminating among genotypes.

Classification of genotypes

Based on the average GY cluster analysis suggested that the 303 lines with GY data in all the eight testing environments could be clustered into 10 groups (Figure 3.2). Cluster validation confirmed that the optimal number of clusters was 10 with Dunn score 0.170. The 10 genotype groups (GGs) had 33, 32, 25, 39, 21, 23, 23, 45, 10 and 52 members (Table 3.5), respectively. The group members were given in the Appendix Table S1. IRRI lines were distributed to all the GGs and dominated in GG7, GG2and GG9. GG8 and G10 had relatively small proportions of IRRI lines. The 10 Chinese lines were distributed to 3 groups. The average GY of the 10 GGs across the eight environments were 665.4g, 657.2g, 763.5g,

622.9g, 721.2g, 689.6g, 738.5g, 676.5g, 790.7g and 679.4g, respectively (Table 3.5). The ranges of average GY of each genotype group in the 8 testing environments were also given in the Table 3.5.

In terms of the origin of the testing lines, the 219 IRRI lines and 30 PhilRice lines distributed in all the 10 genotype groups. IRRI lines were dominant in GG7, GG2, GG9, GG1, GG5, GG3, GG6 and GG4 with percentage ranging from 71.8% to 87.0%. The 10 Chinese lines were distributed into the GG3, GG8 and GG10. GG10 was the largest group and containing 5 of 10 Chinese lines.

Performance plot

Genotype group performance across the three environment groups plot was presented in Figure 3.3. It could be seen that GG9, with about 80% genotypes from IRRI, had the highest GY in E1 and E2 and third lowest in E3. GG7, having the highest percentage of IRRI materials (87.0%), had the second and third highest GY in E1 and E2, while it was the lowest in E3. GG2, with around 81% IRRI lines, had the third highest GY in E1, and the last fourth lowest in E2 and E3, respectively. GG1 showed the second last GY in E3, while it was fourth in E1 and fifth in E2. GG10, the largest group and containing most of the Chinese lines, displayed the highest GY in E3 (JX), while it showed the lowest GY in E1 and third lowest in E2. Similarly GG8, showed the second highest GY in E3, while it was the fifth and second last in E1 and E2, respectively. GG6 displayed the third highest GY in E3, while it was the second lowest in E1 and the fifth last in E2. GG3 displayed the last fourth lowest GY in E1, second highest in E2 and the last fifth in E3. GG4 ranked the last in E2, while it was the medium level in E1 and E3. GG5 displayed the fourth highest GY in E2 and E3, while it was the third last in E1. GG10, GG3 and GG5 showed the most stable GY performance and displayed the highest average GY across the three environment groups. In contrast, GG1, GG2 and GG4 had the lowest average GY across the three environment groups.

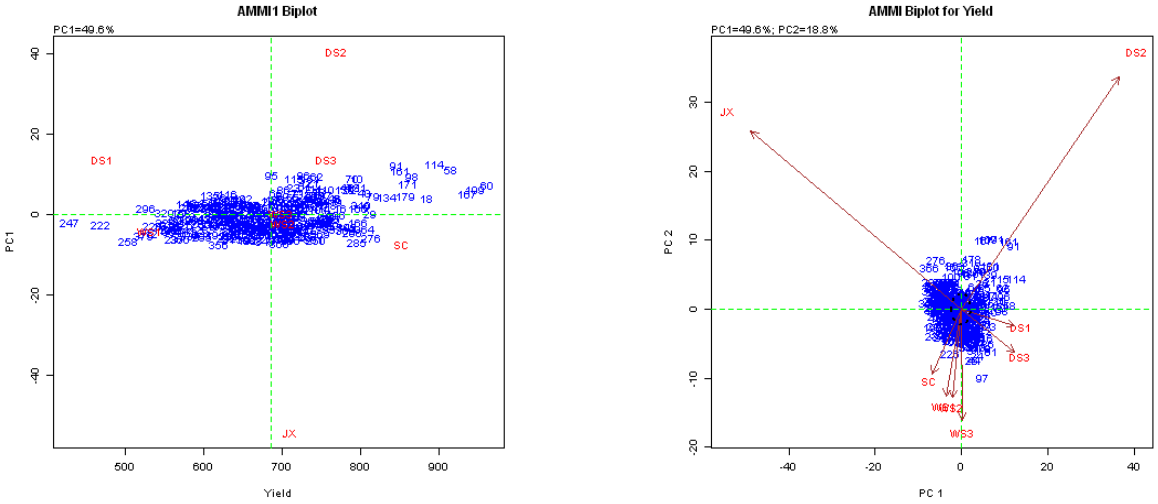
Table 3.5 Basic statistics of 10 genotype groups in eight environments

Env.[a]	Stat.[b]	GG1	GG2	GG3	GG4	GG5	GG6	GG7	GG8	GG9	GG10
DS1	Mean	463.3	438.8	565.3	336.0	483.9	447.9	651.9	424.0	900.3	431.8
	Min	246.5	337.1	302.4	128.6	234.9	231.1	452.6	216.1	715.3	245.2
	Max	669.4	528.7	1237.4	532.1	686.1	645.4	894.9	606.7	1175.4	679.0
DS2	Mean	713.1	733.2	1224.5	533.8	928.9	664.5	1162.8	677.2	904.0	645.0
	Min	491.6	608.1	725.9	324.5	681.2	399.4	727.6	400.8	507.0	363.8
	Max	969.2	1008.6	1787.2	773.1	1350.7	823.6	1681.6	885.0	1375.9	1008.6
DS3	Mean	878.3	689.2	812.7	639.3	884.8	904.1	786.0	645.3	1090.7	687.5
	Min	623.9	425.1	520.6	354.7	686.5	621.1	464.5	378.2	888.8	465.7
	Max	1369.7	864.6	1300.5	818.4	1177.4	1349.3	1184.9	848.5	1281.3	919.2
WS1	Mean	548.6	534.2	558.3	545.1	524.9	513.5	546.0	521.0	554.8	506.0
	Min	386.2	398.0	454.5	381.1	380.1	369.5	431.0	338.9	397.0	308.8
	Max	676.1	660.2	675.5	743.0	692.5	640.7	681.5	686.8	725.1	626.2
WS2	Mean	729.0	720.2	734.1	717.6	686.6	675.7	723.5	673.1	758.1	667.7
	Min	572.3	584.0	609.0	524.2	503.2	418.9	491.5	470.3	706.8	485.0
	Max	807.2	860.0	919.7	888.4	830.6	846.6	859.7	819.2	891.4	873.5
WS3	Mean	756.5	693.1	695.3	719.7	734.8	657.2	729.5	705.8	782.5	623.5
	Min	613.4	531.7	477.9	535.5	528.0	474.1	509.4	442.2	673.9	429.3
	Max	991.1	805.0	836.0	943.7	882.4	805.8	860.5	874.4	900.0	825.8
SC	Mean	807.0	927.1	843.1	762.3	782.4	856.1	897.8	939.7	879.3	836.2
	Min	420.0	764.2	525.9	449.1	559.4	643.0	726.9	744.6	667.4	451.7
	Max	1032.4	1350.3	1102.1	1064.2	985.3	1093.0	1057.5	1160.6	1139.8	1159.4
JX	Mean	427.0	521.7	674.7	729.8	743.5	797.9	410.9	825.9	458.7	1037.4
	Min	220.0	300.0	450.0	405.0	570.0	655.0	200.0	597.5	282.0	760.0
	Max	687.5	732.5	982.5	970.0	1030.0	1012.5	735.0	1085.0	640.0	1372.1
Across.[c]	Min	427.0	438.8	558.3	336.0	483.9	447.9	410.9	424.0	458.7	431.8
	Mean	665.4	657.2	763.5	622.9	721.2	689.6	738.5	676.5	790.7	679.4
	Max	878.3	927.1	1224.5	762.3	928.9	904.1	1162.8	939.7	1090.7	1037.4

[a] Env., Environments

[b] Statistics, Min, minimum; Max, maximum

[c] Across, Across 8 testing environment



a. AMM1 plot b. AMMI2 plot

Figure 3.1 AMMI1 (a) and AMMI2 (b) biplots of 303 rice genotypes tested in eight environments for grain yield.

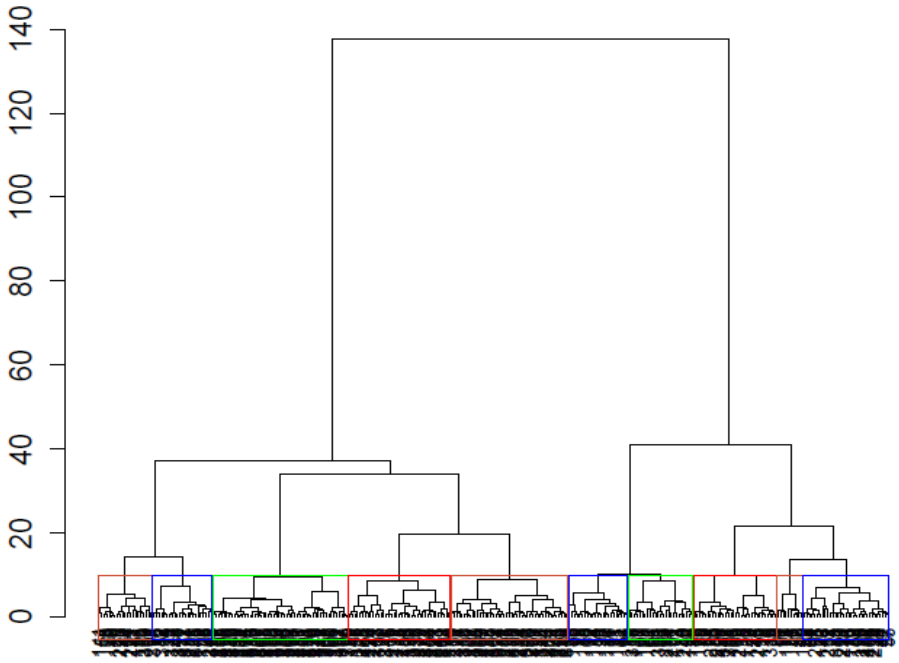


Figure 3.2 Dendrogram of the classification of 303 rice genotypes using Ward's method on environment standardized grain yield.

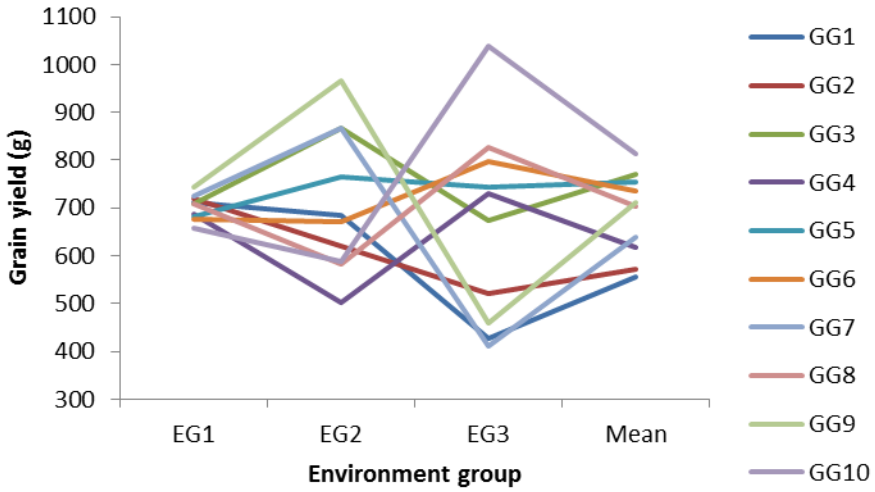


Figure 3.3 The mean grain yield performance plots of genotype groups derived from the classification using Ward's methods against untransformed environment group mean yield

3.4 Discussion

The combined analysis showed highly significant variation for genotype, environment and GEI for GY, indicating genotypes displayed differential expression of yield across the environments. The GEI effect was nearly twice as that of genotype effect. The relative contribution of GEI and genotype to the total variation of GY found in the present study was smaller than those previously found using the INGER irrigated nurseries. This is partially due to the extreme soil and/or weather conditions in some of the testing locations of the INGER nurseries. In the present study, no testing environments were under extreme stress (arguably the no nitrogen treatment can be regarded as abiotic stress). INGER nurseries had many more testing locations in many countries but only a small number of highly selected elite released or pre-released lines. Nevertheless, our results were similar to the results of the studies in rainfed lowland rice (Ouk et al., 2007; Tariku et al., 2013; Wade et al., 1999) and other crops (i. e. wheat, Bertero et al., 2004; Cooper et al., 1996; Canola, Zhang et al., 2013). Thus, it would be very difficult to select for improved lines with broad adaption by conducting selection only in one target environment, ignoring the observed GEI. Better breeding and testing strategy to accommodate the effects of large GEI is required. The following are possible options: (1) Adopting multi-environment testing at the early stages of variety development to allow selecting for general adaptation to be conducted earlier. We regard a program adopting early multi-environment testing allowing early selection for general

adaptation as a potentially efficient way of breeding. An example of this is a collaborative shuttle breeding programme between the Indian Council of Agricultural Research and IRRI. This program has proved to be an effective tool for rice varietal improvement in rainfed lowland ecosystem in eastern India (Mallik et al., 2002). It provided an opportunity for flow of breeding materials of diverse origin among the eastern Indian states to strengthen the breeding programme. The materials developed through this project also served as input to the other breeding programmes for this ecosystem. For those breeders with limited resources, it might be practical to cooperate with peers located different places and thus achieve early multi-environment testing. Across-environment performance must be used as selection criterion when sufficient selection can still be applied. (2) Balancing the number of advanced lines tested in different stages of a multi-stage multi-environment testing scheme. Testing many lines in the first stages and a small number of lines at last stage is not a good option for obtaining reliable GEI information and maximizing genetic gain (Kempton and Fox, 1997). (3) Testing at least 50 lines from a breeding program to include enough genetic variation for the trait of interest. The GEI information obtained from testing a small number of lines from many breeding programs that do not exchange germplasm extensively is less relevant to any breeding program. (4) Subdividing the lowland irrigated ecosystem into more homogeneous TPEs to reduce the effects of GEI if repeatable GEI is identified (Atlin et al., 2000). Critical genotypic characters and soil and weather variables that account for a large proportion of GEI will need to be identified to help defining the TPEs. It should be pointed out that the present study was not aimed at characterizing the TPE of IRRI's irrigated breeding program or the TPE of irrigated lowland ecosystem for *indica* rice in Asia. The number of locations used was few and the trials were conducted only in one year. To characterize the TPE for IRRI's irrigated breeding program a representative sample of IRRI's breeding lines to be tested in many more testing locations across multiple years is needed.

The GEI for GY was partitioned into principal component axis following the AMMI analysis. The first two principal components i.e. IPCA 1 and IPCA 2, accounted for 68.4% of the total variation, were significant and sufficient to explain the GEI. This is in accordance with Gauch and Zobel's (1996) recommendation that the first two IPCAs are usually sufficient. Similarly, Yan and Rajcan (2002) also suggested that most of the interaction occurred in the first few axes. The GSI was the largest source of phenotypic variation for GY and accounted for 22.0% of the total variation. This was inconsistent with Samonte and Hernandez's findings (1990). They found that GSI was significant only in four of their 12 combined analyses for yield and implied that there was no need to conduct stability and

adaptability analysis. The inconsistence between the two results might partially contribute to the different origins of the testing lines. Samonte and Hernandez's data was from lines of four maturity subgroups of the NCT. The testing lines were the promising rice lines selected over a certain number of seasons and over several testing locations nationwide in Philippines. They were more adapted to both of the DS and WS of Philippines. The lines in present study were from many programs including IRRI and PhilRice in Philippines, China and programs in other countries, which were not all subjected to selection under DS and WS. Another more probable reason is that the testing locations in the present study have much bigger differences than those used in Samonte and Hernandez (1990). JX and SC are in the subtropical region while IRRI is in the tropical region. Small GEI observed for yield related traits including DTF, GN, PB, PH, PN, SB, SN, SR, TGW and TN indicated that these traits were relatively more stable among the testing environments. Samonte and Hernandez's findings (1990) also found that GSI and GLI had no significant effect on plant height, tillers and maturity, while the GSLI had significant effect on the three traits in most of the combined analysis.

The DS and WS environments in IRRI were grouped into different groups, indicating that they discriminated the genotypes in different ways. IRRI's irrigated breeding program aims at developing varieties adapted to both of DS and WS. The results of present study showed that the vectors of DS and WS were not in the opposite direction in the biplots (Figure 3.1b), suggesting that it is possible to select genotypes with stable performance across seasons. However, with distinctive and highly repeatable seasonal pattern and different genotype responses to seasons it makes sense that variety development should explore the repeatable GEI caused by season. Much large genetic progress can be made even with the current breeding gene pools by breeding separately for the two seasons. For instance, the average grain yields of the top two genotypes in the DS, entry 60 and 58, were 1264.6g and 1247.9g. In the WS, they ranked the last sixth and sixtieth with average grain yields being only 764.3g and 695.3 g, respectively. Furthermore, it should be pointed out that the majority of the genotypes used in present study are IRRI lines, which have been derived from parental lines selected with stable performance across seasons as the key criterion and as results the interaction between genotype and season was underestimated. It is expected that GSI will be larger if new breeding populations are to be developed using parental materials that have not been selected for adaption to both of the seasons.

The three N treatments in the DS were grouped together to form one group while the three N treatments in the WS were grouped together in another group, indicating that the different N rates used had only a relatively small effect on the relative performance of

genotypes, compared with the season. Previous studies indicated that it is difficult to create useful GEI patterns by use of managed environments in a single location (Cooper et al., 1996). Managed environments can only be useful if they are created by manipulating the key biotic and/or abiotic factors underlying the GEI. Therefore, it may be more appropriate to first investigate the GEI pattern and identify the major reasons for the observed GEI using multi-environment trials to then establish a set of managed environments to measure the GEI.

The IRRI WS environments located close to SC in the biplot. This was consistent with the results of AMMI analysis of the 1994 and 1995 INGER nurseries (INGER 1994a, 1995a). Thus, it seemed that IRRI breeding lines with stable and good performance in the WS could be used in SC (directly as varieties or as parental lines in breeding). Similarly, JX was relatively closer to the IRRI DS in the biplot, suggesting that selection is better to be done in DS in IRRI for use in JX, China. The top 10 varieties recommended for SC, China were Entries 92, 208, 166, 107, 324, 58, 101, 369, 105, and 316. Entries 276, 285, 366, 316, 381, 280, 380, 272, 352, and 349 were best suited for JX, China based on yield performance.

3.5 Conclusion

Using a large number of *indica* genotypes from breeding programs for irrigated ecosystem and 8 testing environments the present study showed that GEI was very important for GY and the genotype-by-season interaction was the major source of GEI. We recommended breeding for different seasons separately to exploit the repeatable GEI caused by seasonal changes. The two testing environments in China were chosen to represent two major distinct rice production environments in China. SC was grouped together with the WS environments of IRRI. JX formed a separate group with more similarity to the DS environments of IRRI. Clearly, great attention should be paid to the relevance of performance at IRRI to their target production environments when IRRI breeding lines are introduced. On the other hand, with a global mandate IRRI's irrigated rice breeding program should expand its testing and selection environments to allow exploiting specific adaption and providing critical and relevant performance information to the developing countries that largely depend on IRRI for new breeding lines.

Chapter 4 Usefulness of the cloned and fine-mapped genes/QTLs for grain yield and related traits in *indica* rice breeding for irrigated ecosystems

Abstract

Many genes/QTLs for grain yield (GY) and yield related traits in rice have been cloned or fine-mapped in the last three decades. A collection of *indica* elite breeding lines and cultivars assembled in IRRI was used to test the usefulness of 39 well characterized yield related genes/QTLs. The population of lines was phenotyped for GY and 10 yield related traits under eight environments of three locations including Jiangxi and Sichuan in China, and six season (2) and nitrogen rate (3) combinations in IRRI and genotyped using 46 markers tightly linked to the 39 target genes/QTLs and 53 SSR markers evenly distributed on the genome. Using the 53 random SSR markers identified two major subpopulations. Association analyses were separately carried out for the whole population and the two subpopulations. All the 39 target genes/QTLs were associated with two or more measured traits including traits not previously reported. *GW6* and *Gn1a* were associated with nine and eight traits, respectively. *Ghd7*, *qSPP7*, *SCM2* and *SPP1* were associated with seven traits. *GIF1* and *Ltn* were associated with six traits. *GS3*, *GW2*, *gw3.1*, *htd1*, *Nop(t)*, *qGY2-1* and *qPH6-1* were associated with five traits. *D10*, *d27*, *DEP2*, *DWL1*, *Gnp4*, *Gw1-1*, *GW3*, *gw5*, *MOC1*, *PAP2*, *qGL7*, *qGL7-2* and *qGN4-1* were associated with four traits, *D88*, *Ghd8*, *GS5*, *Gw1-2*, *IPA1*, *qSH3* and *RPH* were associated with three traits. *ep3*, *gw8.1*, *gw9.1* and *qPDS3* were associated with two traits. A total of 16 genes/QTLs were found to be associated with GY. *GS3*, *GW1-1* and *d27* were associated with GY in two testing environments and the others were only in one environment. Significant gene-by-environment interaction was present for all the studied genes/QTLs. However, GY could not be well predicted using the markers significantly associated with the measured traits or all target markers based on stepwise multiple linear regression analysis. The adjusted coefficient of determination ranged from 0.024 to 0.191 for the final selected models considering the associated markers only and from 0.039 to 0.261 for the final selected models considering all target markers. Nevertheless the known genes might be explicitly utilized in developing more efficient selection criteria for enhancing selection accuracy.

This chapter has been published in the journal of Field Crops Research. Abstract is formatted according to the journal standards

Keywords: Association analysis, Breeding, Grain yield, Marker-assisted selection, QTLs, Rice (*Oryza sativa* L.)

4.1. Introduction

Rice (*Oryza sativa* L.) is one of the most important crops in the world. It has played a central role in human nutrition and been cultured for nearly 10,000 years (Molina et al., 2011). Great progress has occurred in rice production in the last decades due to the adoption of green revolution technology. Paddy rice production has increased from 257 million tons in 1966 to 744.4 million tons in 2014 (FAOSTAT, 2014; Khush, 2005). As world population grows, the global rice demand is estimated to rise from 439 million tons (milled rice) in 2010 to 496 million tons in 2020 and further increase to 555 million tons in 2035 (GRiSP, 2010). At least 50% of the increase of rice production in the past has been due to the adoption of new cultivars. It is expected that breeding will still be one of the key approaches for further increasing in rice productivity. Rice breeders today face the challenge of how to increase productivity by effectively integrating well-established conventional breeding methods with new approaches offered by rapid advances in molecular marker technology and genomics.

For quantitative traits such as yield, many genes of small effects and environmental factors collectively determine the trait performance. The favorable alleles are likely to be spread across more than two lines, therefore requiring the assembly of alleles from different sources to a single inbred line in order to achieve significant improvement (Ye, 2010). Recurrent selection (RS) is a well-established conventional breeding method which is designed to gradually increase the frequency of desirable alleles while maintaining genetic variability for future selections (Hallauer, 1985). RS is a cyclical breeding strategy involving three main steps: selection, evaluation and recombination of the best performing selections. Multiple genotypes are inter-mated to increase the chance of creating novel allelic combinations. Through multiple cycles of recombination between selected genotypes, linkage blocks are broken down and favorable genes are accumulated while the genetic diversity remains (Châtel et al., 2008).

The potential benefits of using molecular markers linked to the genes of interest in breeding programs, which have changed from phenotype-based toward a combination of phenotype- and genotype-based selection, have attracted much attention for more than three decades (Bernardo, 2008; Tester and Langridge, 2010). The success of marker-assisted selection (MAS) for traits of simple inheritance in many crops including rice has motivated

rice breeders to search for QTLs for complex traits, which account for a large proportion of phenotypic variation (major QTLs) (Guo and Ye, 2014). Many yield-related genes/QTLs have been identified and some of them are fine-mapped or cloned (Xing and Zhang, 2010). The effects of the well characterized genes/QTLs were usually tested using specific populations. The use of these well-characterized genes/QTLs in improving yield has started. However, significant improvement of GY in farm environments has yet been reported (Guo and Ye, 2014). In order to successfully deploy MAS, validating the known genes/QTLs in breeding populations and/or in near isogenic lines (NILs) of elite backgrounds is critical. The process of marker validation is required to determine the reliability of a marker to predict phenotype and this points out the advantages of using flanking markers of known genes/QTLs (Collard and Mackill, 2008).

A promising MAS method, known as genomic selection (GS) or genome-wide selection, has been recently introduced for using all trait-affecting genes to improve quantitative traits (Hayes et al., 2013; Meuwissen et al., 2001). GS uses genome-wide markers to predict the breeding (genotypic) values of the selected candidates. Once an accurate prediction model is developed using a reference population with genotyping and phenotypic observations, the model is then used to select genotypes within the selection population with genotyping data only. Simulation and empirical studies in self-pollinated crops, including wheat, barley and oat, have demonstrated the great potential of GS in improving quantitative traits (Jannink, 2010).

The irrigated rice breeding program at the International Rice Research Institute (IRRI) has a mission to develop new rice germplasm that produces high and stable yields across a broad range of environments. IRRI breeding lines have been directly released for production or used as crossing parents in many breeding programs in different countries. IRRI is adopting a new integrated breeding strategy to break yield barrier (GRiSP, 2010). The strategy utilizes RS to quickly pyramid the major QTLs that have been proven useful in the breeding population in the first few selection cycles and maintain the genetic variation contributed by the many minor genes to be explored in later cycles and explores GS for reducing the length of breeding cycles and the cost of expensive phenotyping. A series of studies have been conducted to obtain essential information for designing more efficient mating and selection schemes of this general breeding strategy. As part of this effort, 392 advanced lines and cultivars from many breeding programs in different countries representing the genetic diversity of breeding gene pools for irrigated ecosystems were collected to be used as part of the base population for future breeding at IRRI. This population is being

phenotyped via multi-environment trials (METs) in South-east Asia and genotyped using markers for well-characterized genes/QTLs and genome-wide markers generated by genotyping by sequencing. Data collected so far has been used to analyse GEI for GY (Chapter 3), investigate the factors affecting head rice yield and chalkiness (L. Zhou et al., 2015) and test the usefulness of important known genes/QTLs for grain quality traits (Zhao et al., 2015). In this paper, the usefulness of 39 fine-mapped or cloned genes/QTLs for GY and yield related traits was tested through association analysis. Prediction of GY with markers for these well characterized genes/QTLs was investigated using multiple regression analysis. The results can be used together with information from other studies to optimize the proposed integrated breeding strategy. The information gained will also be valuable to other rice breeding programs willing to exploit the well characterized genes/QTLs for GY and related traits.

4.2. Materials and Methods

Plant materials and phenotyping

Three hundred and ninety two rice lines developed for the irrigated lowland ecosystem were used in this study to achieve a large amount of genetic diversity. The majority of the lines were from IRRI (223). A good number of lines were also from PhilRice (31), CIAT (17), China (13) and Vietnam (11) (Appendix Table S1). Phenotyping was conducted in Jiangxi (JX) and Sichuan (SC) in China and IRRI headquarters (Los Baños, Philippines). The experiments in JX and SC were for one crop season in 2012. At IRRI, the experiment was carried out in the dry season (DS) and wet season (WS) of 2012. Three nitrogen fertilizer application rates, no nitrogen, low (90 kg ha^{-1}) and high (180 kg ha^{-1}), were used to create three artificial environments in the DS, designated as DS1, DS2 and DS3, respectively. Similarly, three nitrogen fertilizer application rates, no nitrogen, low (45 kg ha^{-1}) and high (90 kg ha^{-1}), were used to create three artificial environments in the WS, designated as WS1, WS2 and WS3, respectively. Grian yield (GY) and 10 related traits including grain number per panicle (GN), panicle number (PN), thousand grain weight (TGW), spikelets number per panicle (SN), seed setting rate (SR), number of primary branches per panicle (PB), number of secondary branches per panicle (SB), tiller number (TN), days to flowering (DTF) and plant height (PH), were tested. All 11 traits were measured for the five experiments at IRRI except DS2. In DS2, DTF, PH, PN, GY and TGW were tested. GY, DTF, GN, PH, PN, TGW and

SR were measured in SC, while GY, DTF, GN, PH and SR were measured in JX. Detailed trial description and trait measurement were given in Chapter 3.

Markers and genotyping

Two sets of markers were used. The first set consisted of 53 random SSR markers evenly distributed on the whole genome (Appendix Table S2). The second set consisted of 46 target markers, including SSR, STS and Indel markers which were tightly linked to 39 cloned or fine-mapped genes/QTLs for GY or yield related traits (Table 4.1). Those target markers were initially chosen from the original publications, in which the known genes/QTLs were reported. Even though published information had indicated tight linkage, for implementation in a MAS breeding program, it is very important that further validation be carried out. If there was no polymorphism in our population for any marker, another marker close to the initial marker was chosen. A total of 46 polymorphic markers for the 39 target genes/QTLs were developed and screened for the population.

For DNA extraction the young leaves from a single plant were collected, frozen in liquid nitrogen and stored at -80 °C. DNA isolation was conducted using a cetyltrimethyl ammonium bromide method described by Murray and Thompson (1980). DNA concentration was quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc., Rockland, DE, USA). For PCA we followed the protocol described by Santos et al. (1993) with minor modification. DNA was diluted to 15 ng μl^{-1}. Amplification reactions were carried out in 10µl containing 10 mM Tris-HCL pH 8.3, 50 mM KCL, 1.5 mM $MgCL_2$, 50 µM each of dNTPs, 0.2 µM each of primers, 0.5 U Taq polymerase (Qiangene or homemade) and 10-30 ng of DNA template. PCR reactions were performed using PTC-200 thermal cycler (MJ Research Inc.) or G-Storm thermal cycler. The PCR conditions were as follows: 94 ºC for 5 minutes, followed by 35 cycles of 30 seconds at 94 ºC, 30 seconds at 55 ºC and 45 seconds at 72 ºC with a final extension of 7 minutes at 72 ºC. The PCR amplification products were separated by running on 8% non-denaturing polyacrylamide gels and observed under the AlphaImager (Proteinsimple Inc., California USA) by SYBR® Safe staining method. Each Polyacrylamide gel contained 96 lanes with each holding PCR products from 94 testing lines and two 100 bp DNA ladders (molecular weight ladder) which were located on both side lanes of the gel. Using the molecular weight scoring method, each individual rice plant was scored for each SSR marker.

Statistical analysis

Phenotypic analysis

All trials were separately analyzed by fitting an appropriate spatial model with rows and columns using PBTools (bbi.irri.org) and R (R Core Team, 2015). The best linear unbiased estimations (BLUE) from the best-fit model were used as input data for association analysis. A detailed analysis of phenotypic data has been reported (Chapter 3).

Diversity statistics, population structure and kinship

Polymorphism information content (PIC), gene diversity, number of alleles per locus and heterozygosity were computed using PowerMarker version 3.25 (Liu and Muse, 2005). STRUCTURE 2.3.4 (Falush et al., 2003; Pritchard et al., 2000a) was used to analyze population structure using 53 random markers. The program was run with different values of the number of clusters (k) from one to 10 using the admixture model with correlated allele frequency. Four separate groups of 20 runs were carried out for each K with a burning period of 5,000 followed by 100,000 Markov Chain Monte Carlo (MCMC) repeats. The most likely number of subpopulations was determined using the Delta K method proposed by Evanno et al. (2005). For the optimal k value, the Q matrix (population membership estimates) was extracted from STRUCTURE run. To display the familial relationship, a heatmap of pariwise kinship of the 360 lines was built up in R (R Core Team, 2015).

Association analysis

Rare marker alleles, occurring at a frequency less than 5% among all the genotypes for any marker were excluded. BLUE obtained from the phenotypic analysis were used as phenotypes for each trait and line. Kinship matrix (K) calculation and association analysis were conducted using TASSEL v.2.0.1 (Bradbury et al., 2007). To choose the most appropriate model for association analysis, four models implemented in TASSEL were compared using the 53 random markers and selected traits in selected environments. GY measured in four testing environments including DS2, JX, SC and WS2 and four traits including DTF, GN, GY and TGW measured in WS2 were used to compare the following four models i) naïve model, the general linear model (GLM) without controlling population structure and relatedness; ii) Q model, GLM including Q population structure coefficients as covariates; iii) K model, the mixed linear model (MLM) including the K matrix to account for the genetic relatedness between genotypes; iv) QK model, the MLM including both Q matrix and K matrix as correction for population structure and genetic relatedness between

genotypes. Association analysis was also carried out within the whole population and each of the subpopulations with the most appropriate model. Separate analysis was carried out for all tested traits in each of the eight testing environments and the average environment defined as the average across the testing environments.

To declare significant associations between markers and traits, positive false discovery rate (pFDR; q value) was calculated using the QVALUE in R (Storey, 2002; Storey and Tibshirani, 2003). Since the target markers had been reported to be closely linked to GY or yield related traits and some of these genes have been cloned and validated, a q value cut off of 0.05 was used.

Prediction of GY

Stepwise multiple linear regression (MLR) analysis was used to select an optimal model for predicting GY in each of the testing environments and the average environment. Two sets of markers were used. They were all the 46 target markers and the markers significantly associated with GY or yield related traits identified in each environment using the K model in the present study. To evaluate the relative contribution of each marker and develop the prediction model for GY (Y), the following formula was adopted.

$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_n X_n$

Where Y is the GY, α is the intercept, β_n is the coefficient, X_n is the marker effect.

MLR were conducted using the MASS package in the R (R Core Team, 2015). The stepwise program computed a sequence of MLRs in a stepwise manner. The marker with least AIC value was removed from the regression at each step. The adjusted R^2 of the final model was estimated accordingly.

Table 4.1 Markers, known trait, chromosome location and primers of target genes/QTLs for GY and related traits used for association analysis

Gene/QTL	Marker	Chr.[a]	Pos. (Mb)[b]	Known trait[c]	Forward primer	Backward primer	Remark	Type
SCM2	RM20547	6	27.0	Culm strength	CTCTTCTTCTTCTGTCCGTCTTGG	CCATCTTCATTACCGACCTCTGC	Cloned	SSR
DEP2	RM21964	7	25.3	Dense, erect panicle	AAGACAGCCTTCAAGGGATTTGG	GTACGTGCACCGAGCAGAGC	Cloned	SSR
D88	RM6742	3	5.4	Dwarfness	CCAAGTTATCCAAGCTTCGTTTCG	AGAACGACCTTTCGAGGGAGAGC	Fine mapped	SSR
D10	RM3411	1	31.3	Dwarfness and TN	CGTCCTCCAGATGGTCCAC	ATGGGACTCCCGTACTCCTC	Cloned	SSR
htd1	RM17307	4	27.3	Dwarfness, TN	AGAGCTTGGAGGCACCAATACCG	AGAAAGAACTCCGGCCACCTTCG	Cloned	SSR
DWL1	HL921	3	34.9	Dwarfness, WLT	ATGGCTTCAGACTTCAGAGT	CAAATTAACCTTCAGGCAAG	Fine mapped	STS
DWL1	HL944	3	35.0	Dwarfness, WLT	TGCCAGCCTAGCGAGCCTAA	ATTGCAGCGAGCTACACG	Fine mapped	STS
ep3	STS5803-7	2	9.0	Erect panicle, SB	GAATGGATGGATGGATCGAG	GTAGGGTCCGGCGAGATATG	Cloned	STS
GIF1	RM16942	4	20.4	GIF, TGW	CCAGTACTCTCGCTCCACTCTCC	ATCGCTTTCACGTCACCAAGG	Cloned	SSR
qGL7	RID711	7	28.5	GL, TGW, SN	GCACATGCATGCTAGGACAT	AGCCGGTAAATTTCTTGCAC	Fine mapped	InDel
qGL7-2	Indel1	7	24.6	GL, TGW, SN	CCATAGTAAGACGACCTT	GATATTCTGTCAGCAGTT	Fine mapped	InDel
qGL7-2	RM21945	7	24.9	GL, TGW, SN	CTACACAAGTGAACGCCATCAGG	GTTCTAGGGTGTCCTTTCATGAGC	Fine mapped	SSR
Gn1a	RM3604	1	5.1	GN,GY	CAGGAACCAACCTTCTTCTTGACC	GTCAGACTCCGATCTGGGATGG	Cloned	SSR
Gn1a	RM10316	1	5.3	GN,GY	AAGATCGCTGGGAGATCTGTAGG	GCATGCTAATTAGTCAGCCTTGG	Cloned	SSR
Gnp4	Y48	4	19.6	GN	TCACCATATGGAAGCATCAAG	TATGTGTTTGTTCATGTGCAC	Cloned	SSR
qGN4-1	nkssr04-19	4	31.3	GN, PB, SB,PN	CTGGAATCACAAACCACGAC	GCTACCTCAAGCTCCACGAC	Fine mapped	SSR
qGN4-1	RM3276	4	30.7	GN, PB, SB,PN	TCCGTCTCGACTCTTCCATC	GATGAGACACCACGGACATG	Fine mapped	SSR
GS3	RGS1	3	16.5	GS	TCCACCTGCAGATTTCTTCC	GCTGGTCTTGCACATCTCTCT	Cloned	SSR
GS5	C62	5	3.4	GS	GATTGACTGATAAATTGACAGC	CTAACTCCCATGGAATTAC	Cloned	SSR
GS5	RM574	5	3.4	GS	AAACTAGCCACGGTTTGGTAGGG	AGGGTGGCAGGGATGTAATTTCC	Cloned	SSR
gw5	RMw513	5	5.3	GW, GL	GTATTTGTTTGTCGCATTC	TAGGACCATAGATGTGAGTTA	Cloned	SSR
qGY2-1	RM279	2	2.9	GY	CCTCTCACTCACGTGGACTCTCC	CCTCACCCTAGGCTTTGATATGC	Fine mapped	SSR
Ghd8	RM22483	8	4.3	GY, HD, PH	GACCATGGTGTGAGTGTGACAGG	CAAGTCCTACCTCAACCGCTACC	Cloned	SSR
Ghd7	RM5436	7	9.1	HD, GY, PH, GN	CAAAGGGGGTGTCCTCTATG	GTTGCTCGTCCTACATGTGC	Cloned	SSR
gw9.1	RM24718.CNR111	9	20.9	HD, GY, SN, GN, PH	GACCAACGTGCATGTGACTT	GCTTGCACTAGGGCTCCTT	Fine mapped	SSR
IPA1	RM23422	8	25.2	IPA, TN, GY	GTCGGTCACGAAGTTCAGATCC	TCAGGCAAAGTTGAAGATGGTAGC	Cloned	SSR
Nop(t)	M9	6	24.0	Non-panicle	AATGAATAGATTACCACATGCTA	TGCCTCTTATTTTACTTTTCTTT	Fine mapped	STS

qPH6-1	RM3414	6	2.9	PH	TAGGGCAATTGTGCAAGTGG	TTGGGAATTGGGTAGGACAG	Fine mapped	SSR
qPH6-1	RM19417	6	3.0	PH	CGGCTCCCTTGAATCTTCTGC	GGTGACGCCTTATGAATGTACGC	Fine mapped	SSR
RPH	RM481	7	NA	PH	TAGCTAGCCGATTGAATGGC	CTCCACCTCCTATGTTGTTG	Fine mapped	SSR
d27	RM26985	11	21.4	PH, TN	CACAAGACAACCTTCAATGG	GGCTTAGGAGCGTTTATAGG	Cloned	SSR
qSH3	RM16	3	22.9	Seed shattering	GTGCGCCAGGAGTAGTTGTCTCC	GACGTGTACACATAGCCAAATCATCC	Fine mapped	SSR
PAP2	RM15937	3	30.9	SN	GGAAGAACCTGCGTATCAAGACC	CCACACGGAAGCAGAATTAGCC	Cloned	SSR
PAP2	RM15948	3	31.2	SN	GAGAGTGGGTGAGAAGGATCAGC	ACAGCAGATTTGTTGGCATCTGG	Cloned	SSR
qPDS3	RM14820	3	10.8	SN	AGGTCGTCGATGTCCCTCTCC	AAACCATCGTGGCATCATCTCC	Fine mapped	SSR
qSPP7	RM5499	7	10.0	SN	GGACGAAAGGGTATTTGATTGG	CCTCAAGGTGGTCTCCTTCTCC	Fine mapped	SSR
SPP1	YN27	1	6.5	SN	TACCACTGAACCCACGTGTC	GCTGCCTTTGTTCTCACGTT	Fine mapped	InDel
Gw1-1	RM10376	1	6.3	TGW	TTAGTTTAACCGCACCGTACACC	GGTCGTTGAATTGGTGTCAAGC	Fine mapped	SSR
Gw1-2	RM1344	1	7.0	TGW	CTGCAATCCGAGTAGGAAGC	TGAGCATTCACTCCGATCTG	Fine mapped	SSR
GW3	RM15206	3	16.8	TGW	CATTTCTTTGCCCTCGATCTTTCC	AAGCGCCATAATCCAGGAACC	Fine mapped	SSR
gw3.1	JL109	3	16.6	TGW	TGGAGCTGTGGACTACTGGA;	TCCCTGAGCCTACCTGTCAT	Fine mapped	SSR
GW6	RM20201	6	20.2	TGW	TTAGAGGTAACGGAGGCACAACC	GATGGCTTGAGAGCGTTTGTAGG	Fine mapped	SSR
gw8.1	RM23201.CNR151	8	21.5	TGW, GL	GTTCTTTCCGGTGACGAGAC	CGCTGCAGATGAGCAGATAC	Cloned	SSR
GW2	RM12827	2	8.1	TGW, GW	GCTCTGGCACCGAGATTATTATAGC	GAGAGACTGCGACCTCTGTAGCC	Cloned	SSR
Ltn	ssr0649-23	8	25.3	TN	TCATCCTTAAGACGGTATCACA	CTCCCTCTCCGTTTCATATTC	Fine mapped	SSR
MOC1	RM20373	6	23.9	TN	GACGACGTGTGTTTGACTTCTGC	CCAGTTCCCAACACAAATGAGC	Cloned	SSR

[a] chr., chromosome, the target gene/QTL located

[b] Pos., position, the physical position in Mb

[c] Known trait, the trait which the target gene/QTL controlled; GIF, grain incomplete filling; GL, gain length; GN, grain number per panicle; GS, grain size; GW, grain width; GY, grain yield; HD, heading data; IPA, ideal plant architecture; PB, primary branch number per panicle; PH, plant height; PN, panicle number; SB, secondary branch number per panicle; SN, spikelet number per panicle; SR, seed setting rate; TGW, thousand grain weight; TN, tiller number; WLT, withered leaf tip

4.3. Results

Population structure, genetic diversity and genetic relationship

Population structure analysis indicated that the most likely number of subpopulations was two based on the change of K. Subpopulations 1 and 2 consisted of 155 and 205 lines, respectively (Figure 4.1). Subpopulation 1 was dominated by IRRI materials with one line from AfricaRice and Pakistan, respectively, three lines from Indonesia and four lines from Vietnam. Subpopulation 2 consisted of one line from Pakistan, three lines from AfricaRice, India, Indonesia, respectively, six lines from Bangladesh and Vietnam, respectively, 14 lines from China, 19 lines from CIAT, 25 lines from PhilRice and the remaining 125 lines from IRRI. In summary, subpopulations 1 and 2 had 85.8% and 61.0% lines from IRRI breeding programs, respectively.

PIC, gene diversity, number of alleles per locus and heterozygosity in the whole population and within subpopulations are summarized in Table 4.2. For the whole population, a total of 388 alleles were obtained from the 99 loci including the target and random markers scored for the 360 genotypes, with an average 3.919 alleles per locus varying from two to 11. The average PIC value was 0.3787, ranging from 0.0273 to 0.7826. The average diversity index was 0.4181, ranging from 0.0275 to 0.8079. The average heterozygosity was 0.0058, ranging from 0 to 0.0611, while 65 loci were homozygous. Within subpopulation 1, a total of 325 alleles were detected from the 99 loci, ranging from two to seven alleles per genetic marker, with an average of 3.283 alleles per locus. The corresponding PIC values ranged from 0.0127 to 0.7754, with an average of 0.3399. The genetic diversity at each marker ranged from 0.0128 to 0.8024 with an average of 0.3778. The average heterozygosity was 0.0060, ranging from 0 to 0.0710, while 71 loci were homozygous. For subpopulation 2, a total of 378 alleles were obtained, with an average 3.818 alleles per locus varying from 2 to 8. The corresponding PIC values was 0.3873, varying from 0.0284 to 0.7824. The genetic diversity at each marker ranged from 0.0288 to 0.8077 with an average of 0.4262. The average heterozygosity was 0.0056, ranging from 0 to 0.0732, while 70 loci were homozygous. The genetic diversity within the two subpopulations was consistent with the population structure results. The subpopulation 2 consisted of lines with more diverse origins.

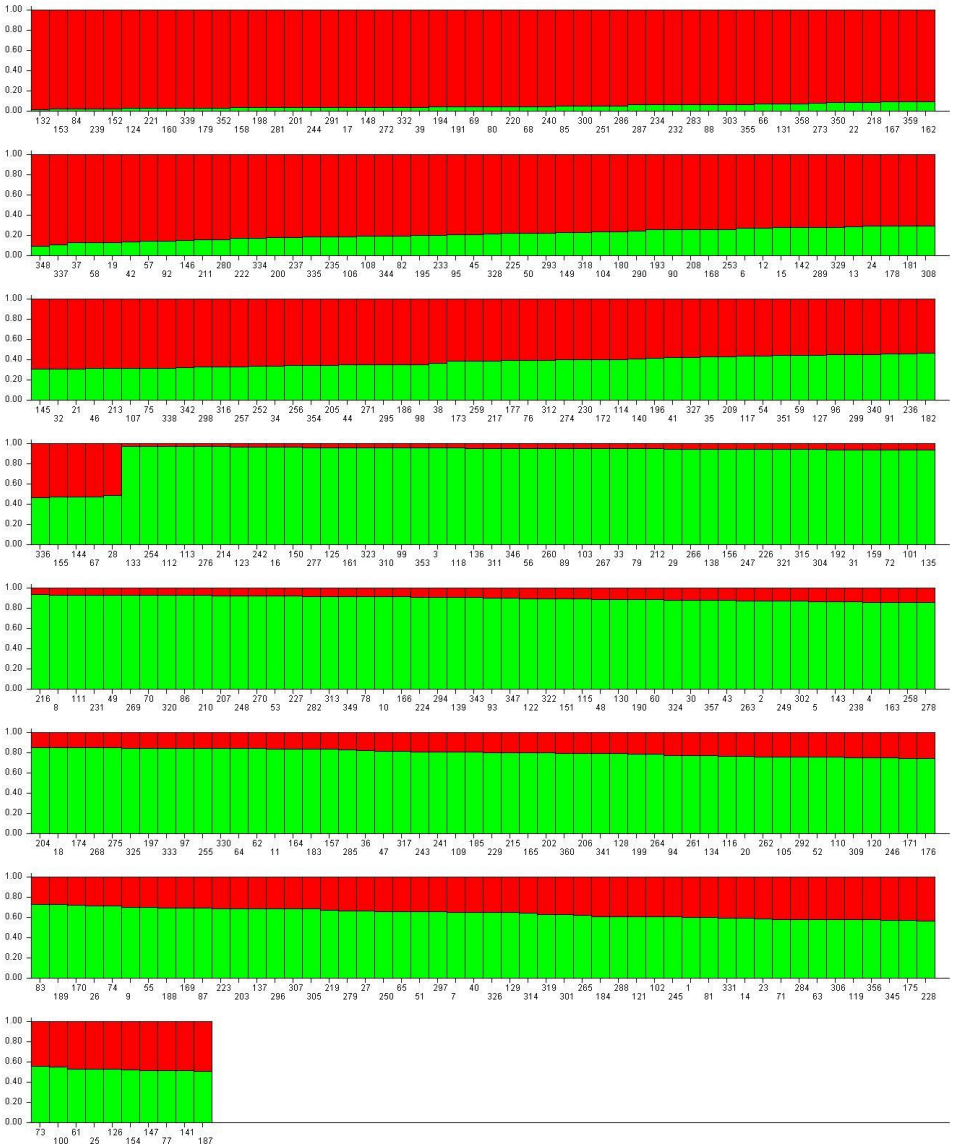


Figure 4.1 Population structure plot of 360 rice genotypes based on 53 unlinked SSR markers obtained using admixture model in software STRUCTURE.

Table 4.2 Summary of molecular diversity and polymorphism information for the whole population and all the subpopulations

Group	No. of lines	Gene diversity	Heterozygosity	PIC	Average allele/locus
Whole population	360	0.4181	0.0058	0.3787	3.919
Subpopulation1	155	0.3778	0.0060	0.3399	3.283
Subpopulation2	205	0.4262	0.0056	0.3873	3.818

The kinship coefficient among the 360 lines ranged from 0 to 1.80, with a mean of 0.49. There were about 2.33% unrelated genotype pairs (kinship = 0), 6.96% distantly related genotype pairs with kinship being lower than 0.10 and 0.03% highly related genotype pairs with kinship being higher than 1.50 (Figure 4.2a). The majority of the genotype pairs had kinship similar to half-sibs (kinship = 0.5). Generally, the relationship within a breeding population was greater than among the breeding populations. The unequal familial relatedness among genotypes could be easily seen in the heatmap (Figure 4.2b).

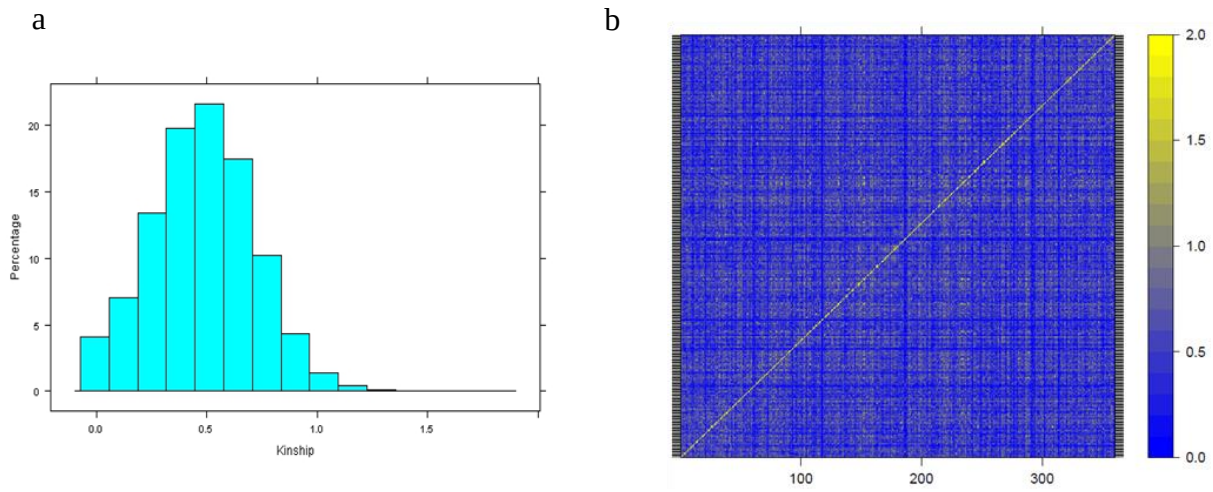


Figure 4.2 Distribution of the pairwise kinship coefficient (a) of 360 lines and heatmap of the estimated pairwise kinship matrix (b) of 360 lines based on the 53 random markers.

Association analysis

Comparison between models

To choose an appropriate model for association analysis, the associations between the 53 random markers and the four traits including DTF, GN, GY and TGW were analysed using four analytical models including the naïve, Q, K and QK models. Model comparison was made using the quantile-quantile (QQ) plot, which is based on the observed and expected *P* values. The ideal distribution of *P* values should follow a uniform distribution with less deviation from the expected *P* values (Kang et al., 2008). Figure 4.3 is the QQ plot for the associations between markers and GY measured in DS2, JX, SC and WS2 using the four models. It showed that *P* values of both naïve and Q models deviated from the Y=X line, especially in JX and WS2. In contrast, the K and QK models showed an improved fit for *P* values and the two models behaved similarly. Figure 4.4 is the QQ plot for the associations between markers and DTF, GN, GY and TGW measured in WS2. The *P* values were inflated for the naive and Q models of. For DTF, GY and TGW, the K and QK models behaved similarly, while for GN, the QK model had a better fit than the K model. The result suggested that mixed linear models using either K alone or both of Q and K sufficiently accounted for structure and familial relatedness among the genotypes in our population. Considering the insignificant difference between the K and QK model and the likelihood that the QK model might be too stringent, only the results from the K model were presented.

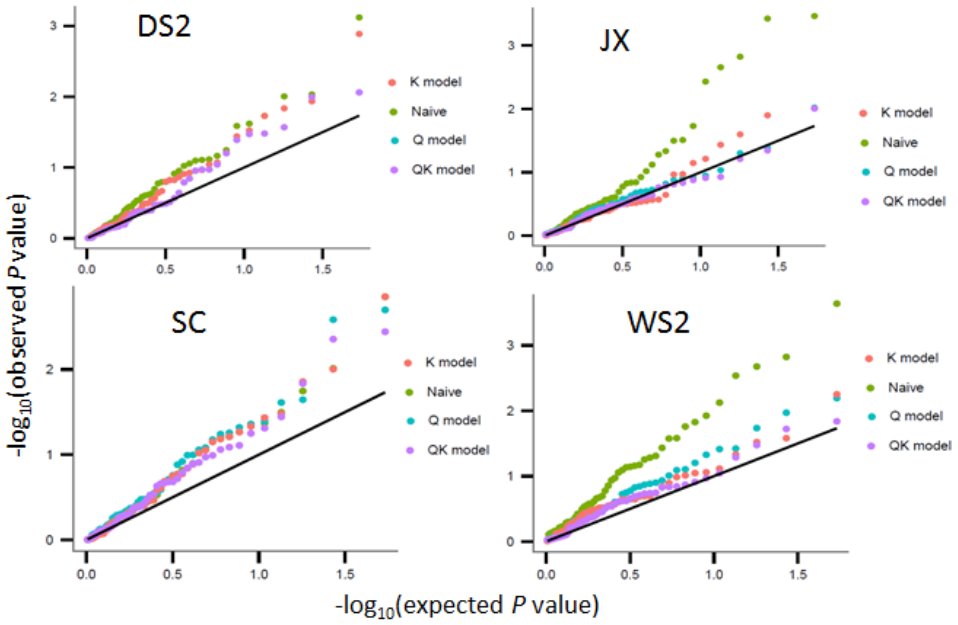


Figure 4.3 Log quantile-quantile (QQ) P value plot for naïve, Q, K and QK models for GY in environments DS2, JX, SC and WS3.

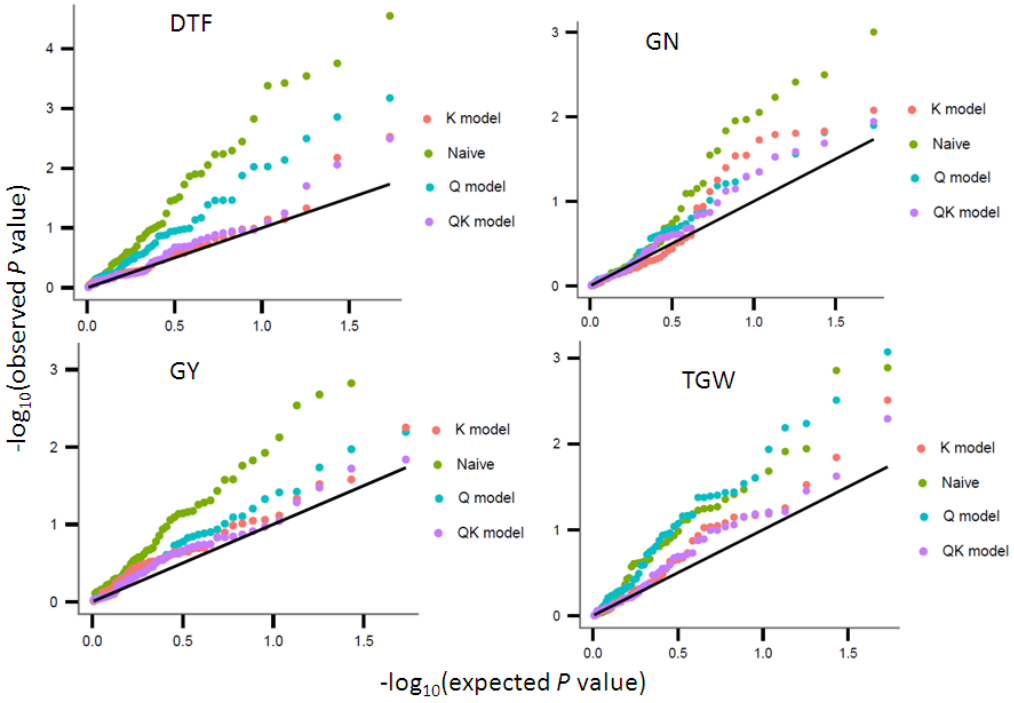


Figure 4.4 Log quantile-quantile (QQ) P value plot for naïve, Q, K and QK models for traits DTF, GN, GY and TGW in environment WS2.

Association between genes/QTLs and traits

Based on the K model applied to the whole population and the two subpopulations, all the 39 genes/QTLs were found to be associated with two or more traits in one or more testing environments. A brief summary of significant associations between the genes/QTLs and the 11 traits is presented in Table 4.3. Gene/QTL-wise, *GW6* was associated with the largest number of measured traits, which was nine, followed by *Gn1a,* which was associated with eight traits. *Ghd7*, *qSPP7*, *SCM2* and *SPP1* were associated with seven traits. *GIF1* and *Ltn* were associated with six traits, respectively. *GS3*, *GW2*, *gw3.1*, *htd1*, *Nop(t)*, *qGY2-1* and *qPH6-1* were associated with five traits, respectively. *D10*, *d27*, *DEP2*, *DWL1*, *Gnp4*, *Gw1-1*, *GW3*, *gw5*, *MOC1*, *PAP2*, *qGL7*, *qGL7-2* and *qGN4-1* were associated with four traits. *D88*, *Ghd8*, *GS5*, *Gw1-2*, *IPA1*, *qSH3* and *RPH* were associated with three traits and *ep3*, *gw8.1*, *gw9.1* and *qPDS3* were associated with two traits (Table 4.3).

All the 39 target genes/QTLs were found to be associated with one or more traits in the average environment (Table 4.3). DS3 had the second highest number (34) of significant gene/QTL trait associations (GTAs), followed by WS1 (20), SC (19) and JX (16). The numbers of significant GTAs found in WS2, WS3, DS1 and DS2 were 14, ten, eight and seven, respectively. The total numbers of significant GTAs were 49 and 44 in the DS and WS, respectively. Nitrogen fertilizer rate had different effects on number of the GTAs in the DS and WS. In the DS, DS3 (high nitrogen rate) had the highest number of GTAs (34), followed by DS1 (no nitrogen) (eight) and DS2 (medium nitrogen level) (seven). In the WS, the highest number of genes/QTLs associated with the measured traits (20) was found under the WS1 (no nitrogen), followed by WS2 (low nitrogen rate) (14) and WS3 (high nitrogen rate) (10).

Table 4.3 Summary of association between target genes/QTLs with GY and yield related traits assessed in eight environments and average environment using K model in the whole population and subpopulations

Gene/QTL	Known trait[a]	DTF	GN	GY	PB	PH	PN	SB	SN	SR	TGW	TN	Sum[b]
D10	Dwarfness and TN	NA	NA	1	1	2	NA	NA	NA	1	NA	NA	4
d27	PH, TN	1	NA	2	1	1	NA	NA	NA	NA	NA	NA	4
D88	Dwarfness	2	NA	NA	1	1	NA	NA	NA	NA	NA	NA	3
DEP2	Dense, erect panicle	1	NA	NA	NA	1	NA	NA	1	1	NA	NA	4
DWL1	Dwarfness, WLT	1	NA	NA	1	1	2	NA	NA	NA	NA	NA	4
ep3	Erect panicle, SB	NA	NA	NA	NA	1	NA	NA	NA	NA	NA	1	2
Ghd7	PH, HD, GY, GN	4	4	NA	NA	1	NA	3	5	2	3	NA	7
Ghd8	PH, HD, GY	1	NA	NA	1	2	NA	NA	NA	NA	NA	NA	3
GIF1	GIF, TGW	5	3	1	1	1	NA	NA	NA	2	NA	NA	6
Gn1a	GN	4	4	1	1	1	1	5	5	NA	NA	NA	8
Gnp4	GN	NA	1	1	1	2	NA	NA	NA	NA	NA	NA	4
GS3	GS	7	NA	3	1	1	NA	NA	NA	NA	5	NA	5
GS5	GS	3	NA	NA	1	1	NA	NA	NA	NA	NA	NA	3
Gw1-1	TGW	NA	3	2	1	2	NA	NA	NA	NA	NA	NA	4
Gw1-2	TGW	NA	1	NA	NA	1	1	NA	NA	NA	NA	NA	3
GW2	TGW, GW	2	2	NA	1	1	NA	NA	1	NA	NA	NA	5
GW3	TGW	1	NA	NA	1	1	NA	NA	NA	NA	3	NA	4
gw3.1	TGW	2	NA	1	1	2	NA	NA	NA	1	NA	NA	5
gw5	GW, GL	1	NA	NA	1	2	NA	NA	NA	NA	2	NA	4
GW6	TGW	1	1	NA	1	1	1	NA	1	1	6	2	9
gw8.1	TGW, GL	NA	NA	NA	1	1	NA	NA	NA	NA	NA	NA	2
gw9.1	HD, GY, SN, GN, PH	NA	NA	NA	1	1	NA	NA	NA	NA	NA	NA	2
htd1	Dwarfness, TN	1	NA	1	1	2	NA	NA	NA	NA	1	NA	5
IPA1	IPA, TN, GY	1	NA	NA	1	1	NA	NA	NA	NA	NA	NA	3
Ltn	TN	5	3	NA	3	2	NA	NA	2	1	NA	NA	6
MOC1	TN	NA	NA	NA	1	1	NA	NA	NA	1	NA	1	4
Nop(t)	Non-panicle	NA	1	1	1	1	NA	NA	2	NA	NA	NA	5
PAP2	SN	NA	NA	1	1	1	NA	NA	NA	2	NA	NA	4
qGL7	GL, TGW, SN	NA	1	1	NA	1	NA	NA	NA	1	NA	NA	4
qGL7-2	GL, TGW, SN	NA	1	1	NA	1	NA	NA	NA	NA	1	NA	4
qGN4-1	GN, PB, SB, PN	NA	NA	1	2	1	NA	NA	NA	NA	1	NA	4
qGY2-1	GY	2	NA	NA	2	2	NA	NA	1	NA	NA	2	5
qPDS3	SN	1	NA	NA	NA	1	NA	NA	NA	NA	NA	NA	2
qPH6-1	PH	2	1	1	NA	2	NA	NA	NA	NA	6	NA	5
qSH3	Seed shattering	1	NA	NA	NA	1	NA	1	NA	NA	NA	NA	3
qSPP7	SN	5	4	1	NA	1	NA	3	5	2	NA	NA	7
RPH	PH	3	NA	NA	1	1	NA	NA	NA	NA	NA	NA	3
SCM2	Culm strength	2	2	NA	1	1	NA	1	4	NA	1	NA	7
SPP1	SN	NA	2	NA	1	2	1	2	3	NA	NA	1	7
Sum[c]	NA	25	16	16	29	39	5	6	11	11	10	5	39
Sum (Whole)[d]	NA	17	13	10	28	6	4	4	8	7	10	1	37
Sum (Sub)[e]	NA	23	11	12	4	39	1	6	10	8	3	5	39

[a] Known trait, the trait which the target gene/QTL controlled, previously reported; [b] Sum of the traits that the gene/QTL associated with; [c] Sum of the genes/QTLs which were associated with the corresponding traits detected in the whole population and the subpopulations; [d] Sum of the genes/QTLs which were associated with the corresponding traits detected in the whole population; [e] Sum of the genes/QTLs which were associated with the corresponding traits detected in the subpopulations.

A total of 16 genes/QTLs were found to be associated with GY using the whole population or the subpopulations (Table 4.4). Ten genes/QTLs were associated with GY using the whole population in one or two testing environments (Table 4.4). *GS3* was associated with GY in two testing environments as well as the average environment. The remaining genes/QTLs were associated with GY in only one testing environment. Separate analysis using the two subpopulations found 12 genes/QTLs associated with GY, in which five were identified in the subpopulation 1 and seven in the subpopulation 2 (Table 4.4). These included *GIF1*, *GS3*, *Gw1-1*, *gw3.1*, *qPH6-1* and *qSPP7* that were also found using the whole population. *d27* was associated with GY in two testing environments in the subpopulation 1, while the remaining genes/QTLs were associated with GY in only one testing environment.

Table 4.4 Association mapping results for GY using K model in the whole population and two subpopulations and target markers with q value less than 0.05

Gene/QTL	Marker	K-whole population[a]		K-subpopulation[b]	
		Environment[c]	q value	Environment (Sub)[d]	q value
D10	RM3411	NA	NA	JX(1)	0.001
d27	RM26985	NA	NA	DS3(1),JX(1)	0.005-0.027
GIF1	RM16942	SC	0.004	SC(2)	0.033
Gn1a	RM10316	NA	NA	DS2(2)	0.037
Gnp4	Y48	SC	0.026	NA	NA
GS3	RGS1	Ave,DS1, DS3	0.0002-0.044	Ave(1),DS1(2), WS2(2),DS1(4)	0.014-0.047
Gw1-1	RM10376	SC	0.004	JX(1)	0.006
gw3.1	JL109	DS1	0.021	DS1(2)	0.014
htd1	RM17307	SC	0.008	NA	NA
Nop(t)	M9	SC	0.029	NA	NA
PAP2	RM15937	NA	NA	JX(1)	0.013
qGL7	RID711	SC	0.013	NA	NA
qGL7-2	Indel1	NA	NA	JX(1)	0.003
qGN4-1	nkssr04.19	NA	NA	DS1(2)	0.031
qPH6-1	RM19417	SC	0.012	NA	NA
qPH6-1	RM3414	SC	0.010	SC(2)	0.033
qSPP7	RM5499	SC	0.010	SC(2)	0.033

[a] K model applied in the whole population

[b] K model applied in the subpopulations

[c] Environment, where the significant MTAs were detected

[d] Environment(Sub),where and which subpopulation the significant MTAs were detected

Twenty five genes/QTLs were found to be associated with DTF using the whole population or each of the subpopulations (Table 4.5). Seventeen genes/QTLs showed significant associations with DTF in one or more testing environments using the whole population (Table 4.5). *GS3* was associated with DTF in five out of the eight testing environments as well as the average environment. *Ltn* and *Ghd7* were associated with DTF in four and three testing environments, respectively. *GS5* was associated with DTF in three testing environments and the average environment. *GIF1* and *GW2* were associated with DTF in two testing environments. *gw3.1* and *qGY2-1* were associated with DTF in one testing environment and the average environment. The remaining genes/QTLs were associated with DTF only in one testing environment or the average environment. When the K model was applied to each of the two subpopulations, a total of 23 genes/QTLs showed significant association with DTF (Table 4.5). All the association were detected only in the subpopulation 2. Eight of the genes/QTLs including *DEP2*, *DWL1*, *Gn1a*, *GW3*, *gw5*, *GW6*, *htd1* and *qPDS3* were not associated with DTF using the whole population. *GS3* was associated with DTF in five testing environments and the average environment. *GIF1* and *qSPP7* were associated with DTF in four testing environments and the average environment. *Ghd7* was associated with DTF in three testing environments and the average environment. *Gn1a* was associated with DTF in four testing environments. *GS5* and *Ltn* were associated with DTF in three testing environments, while *D88* and *qPH6-1* were found to be associated with DTF in two testing environments. *RPH* was associated with DTF in one testing environment and the average environment. The remaining genes/QTLs were associated with DTF in only one testing environment.

All the 39 target genes/QTLs were associated with PH in the whole population or the two subpopulations (Table 4.6). The associations between six genes/QTLs and PH were found in one environment using the whole population, which were *Ghd8*, *Gnp4*, *gw3.1*, *htd1*, *Ltn* and *qPH6-1*(Table 4.6). When association analysis was carried out using the two subpopulations separately, all the target 39 genes/QTLs were found to be associated with PH (Table 4.6). *D10* was significantly associated with PH in the average environment in the subpopulation 2 and in SC in the subpopulation 1, while the remaining genes/QTLs were only detected in the subpopulation 2. *Ghd8*, *Gw1-1*, *gw5*, *qGY2-1* and *qPH6-1* were associated with PH in one testing environment and the average environment, while the remaining genes/QTLs were detected to be associated with PH only in the average environment.

Association analysis using the whole population and each of the two subpopulations found totally 16 genes/QTLs associated with GN (Table 4.7). Thirteen genes/QTLs were found to be associated with GN in one or more testing environments using the whole population (Table 4.7). *Gn1a*, *Ghd7* and *qSPP7* were associated with GN in four out of the eight testing environments. *GIF1* and *Ltn* were associated with GN in two testing environments and the average environment. *SCM2* and *SPP1* were associated with GN in two testing environments. *Gw1-1* and *GW2* were associated with GN in one testing environment and the average environment. The remaining genes/QTLs were associated with GN in one testing environment or the average environment. When the analysis was carried out in the two subpopulations independently, 11 genes/QTLs were found to be associated with GN (Table 4.7). Five genes/QTLs including *Gnp4*, *Gw1-1*, *qGL7*, *qGL7-2* and *qSPP7* were found to be associated with GN in the subpopulation 1 and eight genes/QTLs including *Ghd7*, *GIF1*, *Gw1-1*, *GW2*, *Ltn*, *Nop(t)*, *qPH6-1* and *qSPP7* in the subpopulation 2. Eight of these 11 genes/QTLs were among those found using the whole population except *Gnp4*, *qGL7* and *qGL7-2*. *Gw1-1* and *GIF1* were associated with GN in two testing environments as well as the average environment. *GW2* and *Ltn* were associated with GN in one testing environment and the average environment in the subpopulation 2. The remaining genes/QTLs were associated with GN in only one testing environment.

Table 4.5 Association mapping results for DTF using K model in the whole population and two subpopulations and target markers with q value less than 0.05

Gene/QTL	Marker	K-whole population		K-subpopulation	
		Environment	q value	Environment (Sub)	q value
d27	RM26985	DS2	0.035	NA	NA
D88	RM6742	JX	0.032	JX(2),WS1(2)	0.047-0.048
DEP2	RM21964	NA	NA	WS1(2)	0.037
DWL1	HL944	NA	NA	WS1(2)	0.038
Ghd7	RM5436	Ave,WS1, WS2,WS3	0.001-0.040	Ave(2),WS1(2), WS2(2),WS3(2)	0.005-0.035
Ghd8	RM22483	DS3	0.010	DS3(2)	0.043
GIF1	RM16942	DS2,DS3	0.015-0.017	Ave(2),DS1(2),DS2(2), DS3(2),WS1(2)	0.006-0.036
Gn1a	RM3604	NA	NA	SC(2),WS1(2),WS2(2), WS3(2)	0.018-0.045
GS3	RGS1	DS1,DS2,DS3, WS2,WS3	0.004-0.024	Ave(2),DS1(2),DS2(2), DS3(2),WS1(2),WS2(2)	0.011-0.045
GS5	RM574	DS1,DS2,DS3	0.008-0.041	DS1(2),DS2(2),DS3(2)	0.010-0.024
GW2	RM12827	DS2,DS3	0.010-0.041	DS3(2)	0.043
GW3	RM15206	NA	NA	JX(2)	0.047
gw3.1	JL109	Ave,WS1	0.033-0.040	WS1(2)	0.027
gw5	RMw513	NA	NA	WS1(2)	0.027
GW6	RM20201	NA	NA	WS1(2)	0.048
htd1	RM17307	NA	NA	WS1(2)	0.049
IPA1	RM23422	DS3	0.030	NA	NA
Ltn	ssr0649.23	Ave,SC,WS1, WS2,WS3	0.033-0.050	WS1(2),WS(2),WS3(2)	0.027-0.045
qGY2-1	RM279	Ave,WS1	0.033-0.040	WS1(2)	0.027
qPDS3	RM14820	NA	NA	WS1(2)	0.027
qPH6-1	RM19417	JX	0.045	WS1(2),JX(2)	0.027-0.041
qSH3	RM16	JX	0.008	JX(2)	0.042
qSPP7	RM5499	Ave	0.033	Ave(2),JX(2),WS1(2), WS2(2),WS3(2)	0.035-0.048
RPH	RM481	WS3	0.033	Ave(2),WS1(2)	0.027-0.035
SCM2	RM20547	JX	0.045	WS1(2)	0.027

Table 4.6 Association mapping results for PH using K model in the whole population and two subpopulations and target markers with q value less than 0.05

Gene/QTL	Marker	K-whole population		K-subpopulation	
		Environment	q value	Environment (Sub)	q value
D10	RM3411	NA	NA	Ave(2),SC(2)	0.02-0.036
d27	RM26985	NA	NA	Ave(2)	0.027
D88	RM6742	NA	NA	Ave(2)	0.017
DEP2	RM21964	NA	NA	Ave(2)	0.017
DWL1	HL921	NA	NA	Ave(2)	0.022
DWL1	HL944	NA	NA	Ave(2)	0.033
ep3	STS5803.7	NA	NA	Ave(2)	0.029
Ghd7	RM5436	NA	NA	Ave(2)	0.033
Ghd8	RM22483	WS2	0.004	Ave(2),WS2(2)	0.0127-0.029
GIF1	RM16942	NA	NA	Ave(2)	0.021
Gn1a	RM10316	NA	NA	Ave(2)	0.022
Gn1a	RM3604	NA	NA	Ave(2)	0.022
Gnp4	Y48	SC	0.047	Ave(2)	0.021
GS3	RGS1	NA	NA	Ave(2)	0.033
GS5	C62	NA	NA	Ave(2)	0.029
GS5	RM574	NA	NA	Ave(2)	0.022
Gw1-1	RM10376	NA	NA	Ave(2),JX(2)	0.0004-0.013
Gw1-2	RM1344	NA	NA	Ave(2)	0.021
GW2	RM12827	NA	NA	Ave(2)	0.029
GW3	RM15206	NA	NA	Ave(2)	0.022
gw3.1	JL109	WS2	0.046	Ave(2)	0.013
gw5	RMw513	NA	NA	Ave(2),JX(2)	0.022-0.044
GW6	RM20201	NA	NA	Ave(2)	0.021
gw8.1	RM23201.CNR151	NA	NA	Ave(2)	0.022
gw9.1	RM24718.CNR111	NA	NA	Ave(2)	0.029
htd1	RM17307	SC	0.009	Ave(2)	0.021
IPA1	RM23422	NA	NA	Ave(2)	0.021
Ltn	ssr0649.23	WS1	0.006	Ave(2)	0.021
MOC1	RM20373	NA	NA	Ave(2)	0.029
Nop(t)	M9	NA	NA	Ave(2)	0.021
PAP2	RM15937	NA	NA	Ave(2)	0.027
PAP2	RM15948	NA	NA	Ave(2)	0.029
qGL7	RID711	NA	NA	Ave(2)	0.027
qGL7-2	RM21945	NA	NA	Ave(2)	0.027
qGL7-2	Indel1	NA	NA	Ave(2)	0.029
qGN4-1	RM3276	NA	NA	Ave(2)	0.022
qGN4-1	nkssr04.19	NA	NA	Ave(2)	0.033
qGY2-1	RM279	NA	NA	Ave(2),JX(2)	0.012-0.04
qPDS3	RM14820	NA	NA	Ave(2)	0.021
qPH6-1	RM19417	SC	0.004	Ave(2),SC(2)	0.020-0.029
qPH6-1	RM3414	NA	NA	Ave(2)	0.027
qSH3	RM16	NA	NA	Ave(2)	0.017
qSPP7	RM5499	NA	NA	Ave(2)	0.029
RPH	RM481	NA	NA	Ave(2)	0.017
SCM2	RM20547	NA	NA	Ave(2)	0.013
SPP1	YN27	NA	NA	Ave(2)	0.027

Table 4.7 Association mapping results for GN using K model in the whole population and two subpopulations and target markers with q value less than 0.05

Gene/QTL	Marker	K-whole population		K-subpopulation	
		Environment	q value	Environment (Sub)	q value
Ghd7	RM5436	DS3,SC,WS1,WS2	9.9E-06-0.014	SC(2)	0.0007
GIF1	RM16942	Ave,JX,SC	2.2-E05-0.039	Ave(2),JX(2),SC(2)	4.5-E08-0.049
Gn1a	RM10316	DS3,SC,WS1,WS2	0.009-0.044	NA	NA
Gn1a	RM3604	WS2	0.004	NA	NA
Gnp4	Y48	NA	NA	WS3(1)	0.008
Gw1-1	RM10376	Ave, JX	3.6E-10-4.0E-13	Ave(2),JX(2),SC(2)	1.9E-10-3.7E-10
Gw1-2	RM1344	SC	0.044	NA	NA
GW2	RM12827	Ave, JX	0.0004-0.001	Ave(2),JX(2)	0.014-0.047
GW6	RM20201	SC	0.040	NA	NA
Ltn	ssr0649.23	Ave,DS3,JX	1.9E-19-0.017	Ave(2),JX(2)	1.9E-10-3.7E-10
Nop(t)	M9	SC	0.040	SC(2)	0.015
qGL7	RID711	NA	NA	SC(1)	0.008
qGL7-2	RM21945	NA	NA	SC(1)	0.042
qPH6-1	RM3414	SC	0.002	SC(2)	0.005
qSPP7	RM5499	DS3,SC,WS1,WS2	0.0001-0.017	SC(1),SC(2)	0.0007-0.042
SCM2	RM20547	DS1,SC	0.031-0.045	NA	NA
SPP1	YN27	DS3,WS2	0.006-0.033	NA	NA

Eleven genes/QTLs were associated with SN in the whole population or the subpopulations (Table 4.8). Eight genes/QTLs were found to be associated with SN in one or more testing environments in the whole population (Table 4.8). *Gn1a* was associated with SN in five out of the eight testing environments as well as the average environment. *Ghd7* and *qSPP7* were associated with SN in four testing environments and the average environment. *SCM2* and *SPP1* were associated with SN in the average environment and three and two testing environments, respectively. *Ltn* was associated with SN in one testing environment and the average environment, while *GW2* and *Nop(t)* were in one testing environment. A total of ten genes/QTLs were found to be associated with SN by analysing the two subpopulations independently (Table 4.8), which were found only in the subpopulation 2. Seven of these genes/QTLs except *DEP2*, *GW6* and *qGY2-1* also showed significant association with SN using the whole population. *SCM2* was associated with SN in two testing environments and the average environment. *Ghd7* was associated with SN in two testing environments. *Gn1a*, *Ltn* and *qSPP7* showed significant association with SN in one testing environment and the average environment. The associations between SN and remaining genes/QTLs were significant in only one testing environment.

The fewest significant associations were found between genes/QTLs and PN (Table 4.1). Only five genes/QTLs were found to be associated. Four were found using the whole

population and one using the subpopulation 1. Using the whole population *DWL1* was associated with PN in DS2 and the average environment, *Gn1a* and *Gw1-2* in DS3 and *GW6* in DS2. *SPP1* was associated with PN in WS2 using the subpopulation 1.

Only *GW6* was found to be associated with TN using the whole population (Table 4.2). Analysis in each of the two subpopulations found four additional associated genes/QTLs (Table 4.2). *qGY2-1* was associated with TN in WS1 and the average environment in the subpopulation 2, while *ep3, GW6, MOC1* and *SPP1* were associated with TN in only one testing environment in the subpopulation 1.

Ten genes/QTLs were found to be associated with TGW using the whole population (Table 4.3). *qPH6-1* was associated with TGW in five out of the eight testing environments as well as the average environment. *GS3* and *GW6* showed significant association with TGW in four testing environments and the average environment. *Ghd7* and *GW3* were associated with TGW in two testing environments and the average environment. *gw5* was associated in WS1 and the average environment. *htd1, qGL7-2, GN4-1* and *SCM2* were associated in only one testing environment. Three genes/QTLs were found to be associated with TGW by separate analysis of the two subpopulations (Table 4.3), which were all found only in the subpopulation 2. *GW6* was associated with TGW in three testing environments. *qPH6-1* was associated with TGW in SC and WS3. *GS3* showed significant association with TGW only in WS3.

Table 4.8 Association mapping results for SN using K model in the whole population and two subpopulations and target markers with q value less than 0.05

Gene/QTL	Marker	K-whole panel		K-subpopulation	
		Environment	q value	Environment (Sub)	q value
DEP2	RM21964	NA	NA	DS3(2)	0.028
Ghd7	RM5436	Ave,DS3,SC, WS1,WS2	0.008-0.034	DS2(2),SC(2)	0.028-0.049
Gn1a	RM10316	Ave,DS3,SC, WS1,WS2	0.001-0.031	Ave(2),DS2(2)	0.028-0.045
Gn1a	RM3604	WS2	0.038	NA	NA
GW2	RM12827	WS2	0.038	NA	NA
GW6	RM20201	NA	NA	DS23(2)	0.030
Ltn	ssr0649.23	Ave,DS3	0.022-0.047	Ave(2),DS3(2)	0.028-0.046
Nop(t)	M9	WS1	0.036	Ave(2)	0.046
qGY2-1	RM279	NA	NA	Ave(2)	0.046
qSPP7	RM5499	Ave,DS3,SC, WS1,WS2	0.08-0.034	Ave(2),DS3(2)	0.028-0.046
SCM2	RM20547	Ave,DS1,SC,WS2	0.002-0.045	Ave(2),DS1(2),SC(2)	0.044-0.049
SPP1	YN27	Ave, DS3,WS2	0.009-0.047	DS3(2)	0.0278

Table 4.9 Association mapping results for PN using K model in the whole population and two subpopulations and target markers with q value less than 0.05

Gene/QTL	Marker	K-whole population		K-subpopulation	
		Environment	q value	Environment (Sub)	q value
DWL1	HL921	Ave,DS2	0.045-0.047	NA	NA
Gn1a	RM3604	DS3	0.028	NA	NA
Gw1-2	RM1344	DS3	0.019	NE	NA
GW6	RM20201	WS2	0.007	NA	NA
SPP1	YN27	NA	NA	WS2(1)	0.033

Table 4.10 Association mapping results for TN using K model in the whole population and two subpopulations and target markers with q value less than 0.05

Gene/QTL	Marker	K-whole population		K-subpopulation	
		Environment	q value	Environment (Sub)	q value
ep3	STS5803.7	NA	NA	WS2(1)	0.033
GW6	RM20201	WS2	0.03	WS1(1)	0.004
MOC1	RM20373	NA	NA	WS1(1)	0.042
qGY2-1	RM279	NA	NA	Ave(2),WS1(2)	0.001-0.011
SPP1	YN27	NA	NA	WS2(1)	0.014

Table 4.11 Association mapping results for TGW using K model in the whole population and two subpopulations and target markers with q value less than 0.05

Gene/QTL	Marker	K-whole population		K-subpopulation	
		Environment	q value	Environment (Sub)	q value
Ghd7	RM5436	Ave,DS3,SC	0.030-0.047	NA	NA
GS3	RGS1	Ave,DS1,DS2, DS3,WS3	0.011-0.047	WS3(2)	0.039
GW3	RM15206	Ave,SC,WS2	0.03-0.038	NA	NA
gw5	RMw513	Ave,WS1	0.042-0.048	NA	NA
GW6	RM20201	Ave,DS3,WS1, WS2,WS3	0.003-0.048	SC(2),WS2(2),WS3(2)	0.018-0.026
htd1	RM17307	SC	0.021	NA	NA
qGL7-2	RM21945	SC	0.004	NA	NA
qGN4-1	nkssr04.19	SC	0.05	NA	NA
qPH6-1	RM19417	Ave,DS3,WS1, SC,WS2,WS3	0.00003-0.048	SC(2),WS3(2)	0.007-0.039
qPH6-1	RM3414	Ave,SC	0.001-0.037	SC(2)	0.017
SCM2	RM20547	JX	0.002	NA	NA

A total of 29 genes/QTLs were found to be associated with PB using the whole population or the two subpopulations (Table 4.4). Twenty-eight genes/QTLs were associated with PB using the whole population and four using the subpopulation 2 while no significant association was found using the subpopulation 1. *Ltn* showed significant association with PB in DS3, WS2 and the average environment using the whole population, while the remaining genes/QTLs were associated with PB only in DS3. *Gn1a*, *Ltn*, *qGN4-1* and *qGY2-1*, were found to be associated with PB in the average environment using the subpopulation 2 and *Gn1a* was not found to be associated with PB using the whole population.

Table 4.12 Association mapping results for PB using K model in the whole population and two subpopulations and target markers with q value less than 0.05

Gene/QTL	Marker	K-whole population		K-subpopulation	
		Environment	q value	Environment (Sub)	q value
D10	RM3411	DS3	0.025	NA	NA
d27	RM26985	DS3	0.025	NA	NA
D88	RM6742	DS3	0.024	NA	NA
DWL1	HL921	DS3	0.039	NA	NA
DWL1	HL944	DS3	0.027	NA	NA
Ghd8	RM22483	DS3	0.045	NA	NA
GIF1	RM16942	DS3	0.041	NA	NA
Gn1a	RM3604	NA	NA	Ave(2)	0.014
Gnp4	Y48	DS3	0.024	NA	NA
GS3	RGS1	DS3	0.024	NA	NA
GS5	C62	DS3	0.025	NA	NA
Gw1-1	RM10376	DS3	0.032	NA	NA
GW2	RM12827	DS3	0.024	NA	NA
GW3	RM15206	DS3	0.024	NA	NA
gw3.1	JL109	DS3	0.024	NA	NA
gw5	RMw513	DS3	0.050	NA	NA
GW6	RM20201	DS3	0.025	NA	NA
gw8.1	RM23201.CNR151	DS3	0.024	NA	NA
gw9.1	RM24718.CNR111	DS3	0.030	NA	NA
htd1	RM17307	DS3	0.025	NA	NA
IPA1	RM23422	DS3	0.024	NA	NA
Ltn	ssr0649.23	Ave,DS3,WS2	0.0001-0.036	Ave(2)	0.008
MOC1	RM20373	DS3	0.024	NA	NA
Nop(t)	M9	DS3	0.024	NA	NA
PAP2	RM15937	DS3	0.024	NA	NA
qGN4-1	nkssr04.19	DS3	0.045	NA	NA
qGN4-1	RM3276	DS3	0.025	Ave(2)	0.027
qGY2-1	RM279	DS3	0.024	Ave(2)	0.027
RPH	RM481	DS3	0.039	NA	NA
SCM2	RM20547	DS3	0.025	NA	NA
SPP1	YN27	DS3	0.025	NA	NA

A total of six genes/QTLs were associated with SB. Four genes/QTLs were found to be associated with SB using the whole population (Table 4.5). *Gn1a* was associated with SB in five testing environments as well as the average environment. *Ghd7* and *qSPP7* showed significant associations with SB in two testing environments and the average environment. *SPP1* was associated with SB only in the average environment. Six genes/QTLs were found to be associated with SB by analysing each of the two subpopulations separately, of which *qSH3* and *SPP1* were found in the subpopulation 1 and *Ghd7*, *Gn1a*, *qSPP7* and *SCM2* in the subpopulation 2 (Table 4.5). *Gn1a* was associated with SB in DS1 and the average

environment, while the remaining five genes/QTLs were associated with SB in only one testing environment.

Table 4.13 Association mapping results for SB using K model in the whole population and two subpopulations and target markers with q value less than 0.05

Gene/QTL	Marker	K-whole population		K-subpopulation	
		Environment	q value	Environment (Sub)	q value
Ghd7	RM5436	Ave,DS3,WS1	0.008-0.02	WS1(2)	0.037
Gn1a	RM10316	Ave,DS1,DS3, WS1,WS2	0.0003-0.04	Ave(2),DS1(2)	0.006-0.010
Gn1a	RM3604	WS1	0.049412	NA	NA
qSH3	RM16	NA	NA	WS3(1)	0.013
qSPP7	RM5499	Ave,DS3,WS1	0.01-0.03	WS1(2)	0.037
SCM2	RM20547	NA	NA	DS1(2)	0.046
SPP1	YN27	Ave	0.047085	WS2(1)	0.013

Table 4.14 Association mapping results for SR using K model in the whole population and two subpopulations and target markers with q value less than 0.05

Gene/QTL	Marker	K-whole population		K-subpopulation	
		Environment	q value	Environment (Sub)	q value
D10	RM3411	NA	NA	JX(1)	0.043235
DEP2	RM21964	Ave	0.018	NA	NA
Ghd7	RM5436	Ave,SC	0.018-0.048	Ave(2)	0.043095
GIF1	RM16942	Ave	0.045	Ave(2),SC(2)	0.012-0.043
gw3.1	JL109	Ave	0.039	NA	NA
GW6	RM20201	NA	NA	Ave(1)	0.010028
Ltn	ssr0649.23	NA	NA	DS1(2)	0.046
MOC1	RM20373	NA	NA	Ave(1)	0.032938
PAP2	RM15937	SC	0.030	JX(1),SC(2)	0.012-0.036
qGL7	RID711	SC	0.030	NA	NA
qSPP7	RM5499	Ave	0.009	Ave(2),SC(2)	0.041-0.043

Eleven genes/QTLs were associated with SR (Table 4.6). Seven genes/QTLs associated with SR in the whole population. *Ghd7* was associated with SR in SC and the average environment. The remaining six genes/QTLs were associated with SR in only one testing environment or the average environment. Eight significant associations between the target genes/QTLs and the SR were found by analyzing the two subpopulations separately (Table 4.6). *D10*, *GW6* and *MOC1* were associated with SR in the subpopulation 1 while *Ghd7* and *Ltn* in the subpopulation 2. *GIF1* and *qSPP7* were associated with SR in one testing environment and the average environment in the subpopulation 2. *PAP2* was

associated with SR in JX using the subpopulation 1 and SC in the subpopulation 2. The remaining genes/QTLs were associated with SR in one testing environment or the average environment.

Prediction of GY using markers for known genes

Two prediction models were developed to predict GY using stepwise MLR for each of the eight testing environments and the average environment. One was developed using significant markers associated with any of the measured traits in any of the environments (associated marker model) and the other using all the markers for the 39 target genes/QTLs (all marker model). Table 4.7 lists the selected markers and adjusted R^2 in the final models for each environment. In the average environment, all 39 genes/QTLs were associated with at least one of the 11 tested traits, thus there was only one model. Sixteen markers were included in the model and captured 15.8% of the phenotypic variation. For the three environments in the DS, the final associated markers models for DS1, DS2 and DS3 included two out of eight, three out of seven and 14 out of 38 markers and explained 7.0%, 2.4% and 19.1% of the total phenotypic variation, respectively. The final all marker models for DS1, DS2 and DS3 consisted of 12, six and 17 markers and captured 15.1%, 3.9% and 23.0% of the phenotypic variation, respectively. There were two and ten common markers between the two final models for DS1 and DS3, respectively. For DS2, there was no common marker between the two final models. For JX and SC, three out of 17 and 12 out of 22 markers were included in the associated marker models, which explained 6.3% and 14.9% of the total phenotypic variation, respectively. Four and 17 markers were selected in the all marker models and explained 11.7% and 19.3% of the total variation, respectively. There were two common markers between the two models for JX and seven common markers between the two models for SC. The final associated marker models for WS1, WS2 and WS3 included six out of 21, eight out of 15 and five out of ten markers and explained 10.6%, 11.6% and 9.7% of the total variation, respectively. The final combined marker models for WS1, WS2 and WS3 had 17, 21 and 18 markers and captured 22.9%, 26.1% and 22.5% of the phenotypic variation, respectively. The number of common markers between the two final models for WS1, WS2 and WS3 were four, six and two, respectively.

Table 4.15 The final model for predicting GY in each environment resulted from the multiple linear regressions, R^2 and adjusted R^2

Environment	Marker set[a]	Selected markers[b]	R^2	Adjusted R^2
Ave	Associated	RGS1,RM20547,ssr0649.23,RMw513,STS5803.7,Y48,HL944,RM16, RM14820,YN27,RM279,M9,RM3604,RM12827,RM20373,RM21964	0.247	0.158
	All	RGS1,RM20547,ssr0649.23,RMw513,STS5803.7,Y48,HL944,RM16, RM14820,YN27,RM279,M9,RM3604,RM12827,RM20373,RM21964	0.247	0.158
DS1	Associated	RGS1,nkssr04.19	0.089	0.070
	All	RGS1,nkssr04.19,RM14820,RM21964,HL944,RM3411,RM19417,M9, RM26985,RM20373,C62,RM574	0.211	0.151
DS2	Associated	RM12827,RM10316,HL921	0.039	0.024
	All	JL109,RM15937,RGS1,RM23422,Y48,RM22483	0.072	0.039
DS3	Associated	RM10316,RM5436,RM3411,HL944,Y48,JL109,RM20201,RM20373,RM279, RM20547,RM574,C62,RMw513,M9	0.272	0.191
	All	RM10316,RM5436,RM3411,HL944,Y48,JL109,RM20201,RM20373,RM279, RM20547,RM19417,RM14820,RM21945,STS5803.7,Idel1,RM15206,RM3414	0.325	0.230
JX	Associated	RM20547,RM3411,RM279	0.092	0.063
	All	RM20547,RM3414,RM21945,RM20373	0.175	0.117
SC	Associated	RM10376,RID711,RM20547,M9,RM19417,RM15937,RM21945,RM17307,RM3411, RM16942,RM10316,Indel1	0.208	0.149
	All	RM10376,RID711,RM20547,M9,RM19417,RM15937,RM21945,RM5499,RGS1, RM24718.CNR111,RM14820,RM3276,YN27,RM26985,RM3604,RM21964,RM15948	0.277	0.193
WS1	Associated	RMw513,ssr0649.23,M9,RM6742,RM5436,RGS1,	0.151	0.106
	All	RMw513,ssr0649.23,M9,RM6742,RM1344,RM3276,Y48,RM21964,RM3411,YN27, RM279,nkssr04.19,RM19417,HL921,RM26985,RM15206,RM574	0.338	0.229
WS2	Associated	RM5436,ssr0649.23,RM10316,RM20201,RM15206,RGS1,RM19417,RM3604	0.158	0.116
	All	RM5436,ssr0649.23,RM10316,RM20201,RM15206,RGS1,HL921,RM15937,RM17307, RM21964,RM24718.CNR111,RM481,RM574,RM6742,RMw513,Y48,YN27,RM279, RM10376,RM3276,RM23201.CNR151	0.385	0.261
WS3	Associated	RM19417,RM20201,RM3604,RGS1,RM16,	0.119	0.097
	All	RM19417,RM20201,RM6742,RM3411,RM15948,RM23201.CNR151,RM14820,RM3276, RM279,nkssr04.19,RM481,HL921,RM10316,RM20547,RM3414,RM17307,RM20373,C62	0.337	0.225

[a] marker set used for regression. All, all the 46 markers related to the known genes. Associated, detected to be associated with GY or yield related traits in this study; [b] markers selected and included in the final model.

4.4. Discussion

Genetic diversity can be further increased by introducing Chinese breeding lines

One of the objectives of the newly designed irrigated breeding program at IRRI is to increase the genetic diversity of its breeding populations by incorporating elite lines from other large breeding programs. The genetic diversity and PIC value of the present population were 0.4181 and 0.3787, respectively. The PIC value was close to and comparable with Huang et al.'s (2015) results in terms of the diversity level and the population size. They found that the PIC value of 299 inbred *indica* rice varieties cultivated in Guangdong province of China assessed using 34 SSR markers was 0.38. The genetic diversity and PIC value in our study were lower than those of the populations of 83 global *indica* lines (0.694 and 0.665) and 495 Chinese *indica* lines (0.623 and 0.595) assessed by Wang et al. (2014) with 84 SSR.

The average allele per locus was 3.92 in the present population. This was slightly higher than that of Huang et al.'s study (2015), in which the average allele per locus was 3.3 for 34 SSR markers. The average allele number was also higher than that of the population comprised of 328 *indica* accessions mostly collected in China studied using 100 SSR (Jin et al., 2010), in which the average allele per locus ranged from 2.60 to 3.36 among the 6 subpopulations. The average number of alleles per locus in our study was much smaller than Wang et al.'s (2014) results, which reported that the average allele per locus for the 84 SSR in the populations of 83 global lines and 495 Chinese *indica* lines were 10.5 and 13.0, respectively. Considering that our population consisted of lines mainly from IRRI it seemed that genetic diversity can be greatly increased by introducing Chinese breeding lines. The few Chinese breeding lines included in the present study could not represent the genetic diversity of Chinese breeding lines.

Careful association analysis is important

The usefulness of 39 well-characterized genes/QTLs was tested using association analysis in this study. It is well-known that population structure and unequal relatedness among individuals greatly affect the results of association analysis and must be carefully taken into account (Pritchard and Rosenberg, 1999; Yu et al., 2006). The 360 *indica* lines of the present study were divided into two subpopulations using a model based analysis implemented in STRUCTURE (Falush et al., 2003; Pritchard et al., 2000a). The kinship among the 360 lines was unequal. The estimated kinship coefficient ranged from 0 to 1.80 with an average 0.49. Distantly related genotype pairs accounted for 6.96% and highly related genotype pairs

accounted for only 0.03%. The majority of the genotype pairs had kinship similar to half-sibs. The association results from models with and without population structure and relationship among individuals were significantly different (Figures 4.3 and 4.4). The naïve model identified the highest number of marker trait associations, in which there must be a lot of false positives as many of the observed *P* values significantly deviated from the expected *P* values in the QQ plot. The Q model fitted the expected *P* values better than the naïve model but much worse than the QK and K models, indicating that controlling only for the population structure was useful but not enough to reduce the spurious associations. The K model and QK model performed similarly for most traits and in most testing environments, suggesting that controlling the unequal familial relatedness could sufficiently control the spurious associations in the present population. We also found that some genes/QTLs were only detectable in one of the two subpopulations using the K model, although the subpopulation was much smaller than the whole population.

Usefulness of the tested genes/QTLs

All the studied genes/QTLs were associated with at least two of the 11 measured traits in one of the eight testing environments or the average environment. The numbers of genes/QTLs associated with GY, DTF, PH, GN, SN, PN, TN, TGW, PB, SB and SR were 16, 25, 39, 16, 11, five, five, ten, 29, six and 11, respectively (Tables 4.4-4.14). These genes/QTLs could thus be used in improving the corresponding target traits in breeding for different objectives. A few genes/QTLs were associated with trait(s) in more than three testing environments. *GS3* was associated with DTF in six testing environments and the average environment. *GIF1*, *qSPP7* and *Ltn* were associated with DTF in four testing environments as well as the average environment. *Gn1a* associated with DTF in four testing environments. *Ghd7* was associated with DTF in three testing environment and the average environment. *GS5* was associated with DTF in three testing environments. *Ghd7*, *Gn1a* and *qSPP7* were associated with GN in four testing environment. *qPH6-1* and *GW6* were associated with TGW in five testing environments and the average environment. *Gn1a* was associated with SB in four testing environment as well as the average environment. *Ghd7*, *Gn1a* and *qSPP7* were associated with SN in four testing environments and the average environment, while *SCM2* in three testing environments and the average environment. These genes/QTLs could be regarded as the important ones for the corresponding traits. The following observed characteristics need to be carefully considered when these genes/QTLs are explored in breeding.

Firstly, for all genes/QTLs environment had large effects and significant gene-by-environment interaction was present. The total number of genes/QTLs associated with the measured traits varied greatly across the testing environments ranging from seven in DS2 to 34 in DS3. Both of season and nitrogen rates affected the number of GTAs, although season had less effect than nitrogen level. The numbers of GTAs found in the DS and WS were 49 and 44, respectively. Nitrogen rate had different effects on the GTAs in the DS and WS. In the DS, the numbers of significant genes/QTLs were eight, seven and 34 in DS1, DS2 and DS3, respectively while in the WS they were 20, 14 and ten in WS1, WS2 and WS3, respectively.

The importance of GEI was clear, since no gene/QTL was significantly associated with a trait in all the eight testing environments and that the associated genes/QTLs changed from environment to environment for all traits. Some of the genes/QTLs even had effects in opposite directions in different environments. One allele of *Gn1a* had opposite effect on DTF in WS2 compared with in other three environments including SC, WS1 and WS3. Similarly, one allele of *GIF* and *Ltn* had opposite effect on DTF in DS3 and WS3. One allele of *qSPP7* had opposite effect on DTF in JX and WS3 compared to in WS1 and WS2 (Data was not given).

Secondly, all the 39 genes/QTLs were found to be associated with two or more traits in the present study. *GW6* was associated with nine out of the 11 traits excepting GY and SB. *Gn1a* was associated with eight traits excepting SR, TGW and TN. *Ghd7*, *qSPP7*, *SCM2* and *SPP1* associated with seven traits. *GIF1* and *Ltn* were associated with six traits. *GS3*, *GW2*, *gw3.1*, *htd1*, *Nop(t)*, *qGY2-1* and *qPH6-1* were associated with five traits. For the remaining 24 genes/QTLs, 13, seven and four were found to be associated with four, three and two traits, respectively (Table 4.3). Seventeen of the 39 target genes/QTLs were known to affect more than one trait previously (Table 4.1). The underlying biology of pleiotropy might be that a gene was involved in multiple pathways. For instance, *Ghd7* was involved in multiple pathways for traits contributable to yield, such as phytohormone synthesis pathway and photoperiod pathway (Wei et al., 2010; Xue et al., 2008) and biotic and abiotic stresses (Weng et al., 2014). On the other hand, it is possible that genetic strongly correlated agronomic traits are controlled by the same biological processes.

Many of the tested known genes/QTLs were reported to be associated with only one trait in previous reports. This may simply due to the limited number of traits measured. In the present study, we tested 11 yield related traits and found that all the target genes/QTLs were multiple-trait associated. Therefore, it is important for MAS to investigate the effects of the

target genes on other important agronomic traits. This is particularly important when genes/QTLs with limited traits information are used.

Thirdly, all the genes/QTLs were associated with traits unreported previously (Table 4.3). For instance, the grain weight controlling gene, *GW6* (Guo et al., 2009), was associated with TGW and other eight traits except GY and SB. *Gn1a* regulating grain number and enhancing GY (Ashikari et al., 2005) was found to be associated with GN, GY and other six traits including DTF, PB, PH, PN, SB and SN. *Ghd7*, an important regulator of PH, HD, GN and GY (Lu et al., 2012; Xue et al., 2008), was found to be associated with PH, DTF and GN. It was also associated with SB, SN, SR and TGW in one or more environments. This could be partially explained by the fact that most of the gene/QTL fine mapping and cloning studies usually focused on only one or few traits and whether the target gene/QTL affects other traits was not studied. The present study evaluated GY and ten related traits and provided a chance to identify associations between genes/QTLs with traits not measured previously, which furthered our understanding of agronomic importance of the known genes/QTLs.

It is also worthy of pointing out that validation of genes/QTLs before implement in MAS is extremely important. To date, hundreds of studies and reviews have reported QTLs for grain yield and related traits in rice (Xing and Zhang, 2010; Zuo and Li, 2013). However, MAS has contributed very little to the release of varieties with enhanced yield performance. One of the reasons is that QTL mapping and cultivar development are currently separate processes, which often involve different types of parents and populations. This not only increases the time required for cultivar development, but also reduces the probability of successfully utilising QTL information to create a superior variety. Validation of markers linked to QTLs and their effectiveness in postulation and selection in the target phenotype in independent populations and different genetic backgrounds is crucial (Bian et al., 2013; Kim et al., 2016; Liu et al., 2013). Once tightly linked markers that reliably predict a trait phenotype have been identified, they may be used for MAS (Collard and Mackill, 2008).

GY cannot be well predicted using markers for known genes

Using the markers for genes/QTLs associated with yield and related traits to predict GY under each environment captured from 2.4% (DS2) to 19.1% (DS3) of the phenotypic variation. The number of selected markers in the final prediction model varied from three (DS2) to 14 (DS3). The percentage of variation captured by markers in the final models derived from all the target markers varied from 3.9% (DS2) to 26.1% (WS2). As expected, using significant markers identified by association analysis was worse than using all markers.

Clearly, the predictions could not be used to replace phenotypic selection based on GY. Although it was often claimed that a single cloned gene could increase GY sustainably, none of the 39 tested genes/QTLs has enough power to predict GY well in the present study. Nevertheless, it is possible to develop more efficient selection criterion and/or schemes by explicitly incorporating the known genes with large effects. Lines such as entry 369 containing 12 favourable alleles among the testing genes/QTLs could be used in a breeding program. There were 57 lines containing 11 favourable alleles which could be used for yield improvement in a breeding program.

GY prediction based on markers could be further improved by using high density and high throughput genotyping readily available in rice. One option is to identify more MTAs through genome-wide association studies (GWAS). The other option is to exploit GS that utilizes markers in linkage disequilibrium (LD) with all genes affecting GY and captures interactions between genes (Desta and Ortiz, 2014; Nakaya and Isobe, 2012). GS has been applied in rice to predict the breeding values of GY (Spindel et al., 2015). The prediction accuracies ranged from 0.31 to 0.34 for GY and outperformed prediction using MLR and pedigree records alone. GS could be an effective tool to increase the selection efficiency, especially for those traits with complex genetic architecture.

4.5. Conclusion

The usefulness of 39 genes/QTLs were tested using association analysis in a population of wide range of varieties or elite breeding lines, which were chosen to represent the genetic diversity of some *indica* breeding gene pools for irrigated ecosystems (Appendix Table S1). Most of the lines are being used as parental lines in IRRI's irrigated breeding programs including pedigree breeding and RS. All the studied genes/QTLs were associated with at least two of the 11 measured traits in one or more of the eight testing environments or the average environment. For GY, the number of significant genes/QTLs varied from 0 to eight across testing environments and the average environment. It was found that significant GEI was present for all genes, most of the GTAs were not reported previously and all of the genes/QTLs were associated with at least two tested traits. GY could not be well predicted using markers for these known genes. New MTAs need to be identified using GWAS with high density markers and advanced statistical approaches. GS models based on large number of markers for genes affecting GY and captures small QTL effects and interactions between

genes should be exploited. The well characterized genes/QTLs with large effects can be used explicitly in developing GS selection criteria to increase selection efficiency.

Chapter 5 Genome-wide association study of grain yield and related traits using a collection of advanced *indica* rice breeding lines for irrigated ecosystems

Abstract

Genome-wide association study (GWAS) is an effective approach for the identification of marker-trait associations (MTAs) using more diverse germplasm not constrained by specific crossing. This study used a collection of 327 advanced *indica* breeding lines and varieties genotyped using a genotyping-by-sequencing method and phenotyped in eight environments to identify MTAs for grain yield (GY) and 10 related traits. The current panel of *indica* breeding lines were divided into two subpopulations by three analytical methods including STRUCTURE, principal component analysis (PCA) and neighbour joining (NJ) tree. STRUCTURE and PCA gave the same classification of genotypes while the NJ tree results are slight difference in assigning genotypes to subpopulation from STRCUTURE and PCA. Wide variation was observed for all the 11 traits in the whole panel and the two subpopulations inferred by STRUCTURE analysis. Linkage disequilibrium (LD) analysis showed that LD decay varied across the chromosomes and average pairwise squared correlation coefficient (r^2) dropped to half of its maximum value (r^2=0.25) within a physical distance of 200 kb in the whole population. A MLM model controlling both population structure and cryptic relatedness was chosen to identify MTAs for all the 11 tested traits. A total of 452 MTAs that were delineated into 43 QTLs were identified for all traits but PB, SB, and SR with 39 QTLs being not reported before. Three QTLs on chromosome 6, 9 and 12 were identified for GY but only in DS2. The numbers of QTLs identified for the remaining traits varied from two to 26. Most of the detected QTLs were found in only one environment. Four QTLs were located in the regions containing genes/QTLs previously identified for other related traits. The effects of identified QTLs were relatively small with the highest percentage of phenotypic variance explained by a single QTL being 9.6%. The identified QTLs are directly relevant to and can be more effectively used in breeding programs.

***Keywords*:** Rice, Yield-related traits, QTL, Genome-wide association studies (GWAS), Genotype-by-sequencing (GBS), SNP

This chapter has been accepted by the journal of Field Crops Research. Abstract is formatted according to the journal standards

5.1 Introduction

Rice (*Oryza sativa* L.) is extensively cultivated on every continent in more than 100 countries and consumed by more than half of the world's population (Juliano, 1993). As world population grows, the global rice demand is estimated to rise from 439 million tons (milled rice) in 2010 to 496 million tons in 2020 and further increase to 555 million tons in 2035 (GRiSP, 2010). Thus a significant increase in rice production is needed for the future of food security.

The Green Revolution in the 1960's greatly increased rice production with at least 50% of the increase being due to adoption of new cultivars. However, increased production potential of modern rice cultivars has stagnated (Nguyen and Ferrero, 2006) partially due to its narrow genetic base resulting from narrow breeding populations and the population bottleneck that occurred during domestication. How to break the yield barrier is the major challenge for rice breeders (GRiSP, 2010; Ye et al., 2013). On the other hand, as a model crop species for plant molecular biology and genomics, rice has more accumulated molecular and genomic information than most other crops. The utilization of this genetic information offers the rice breeding community a range of modern tools and methods for addressing this challenge. An integrated strategy was proposed by the International Rice Research Institute (IRRI) to increase breeding efficiency by effectively and efficiently utilizing well proven conventional breeding methods, new techniques and methods enabled by modern molecular biology and genomics and advanced methods in experimental design and data analysis (GRiSP, 2010; Ye et al., 2013). This strategy utilizes (marker-assisted) recurrent selection to quickly pyramid the major genes/QTLs in the first few selection cycles and maintain genetic variation contributed by many minor genes to be explored in later cycles, explores genomic selection (GS) for reducing the breeding cycles and the costs of phenotyping and adopts advanced experimental design and data analysis methods to improve heritability (Ye et al., 2013).

To implement this strategy 392 advanced lines and cultivars from many breeding programs in different countries representing the genetic diversity of breeding gene pools for irrigated ecosystems were collected to be used as part of the base breeding population. This population has been phenotyped via multi-environment trials (METs) in South-east Asia and genotyped using markers for well-characterized genes/QTLs and genome-wide markers to obtain essential information for designing more efficient mating and selection schemes of this general breeding strategy. Phenotypic analysis and basic genotypic analysis were used to

remove erroneous lines and lines unadapted to the targeted tropical and subtropical environments and to investigate genotype-by-environment interaction (GEI) for GY (Chapter 3). Thirty-nine cloned or fine-mapped genes/QTLs for GY and related traits were tested for their usefulness through association analysis (Chapter 4). Significant gene-by-environment interaction was found for all of the genes/QTLs and GY could not be well predicted using the markers significantly associated with the measured traits. More marker-trait associations (MTAs) need to be identified.

Considering that the majority of the well characterized genes/QTLs were identified using biparental populations of contrasting phenotypes it is not unexpected that their transferability to breeding populations is low. An alternative method for the detection of MTAs is association mapping (AM). AM identifies MTAs caused by linkage disequilibrium (LD), which is the non-random association of alleles at separate loci. AM utilizes ancestral recombination events to identify MTAs and provides comparatively higher mapping resolution than the biparental linkage analysis (Zhu et al., 2008). AM has been successfully employed in rice to identify MTAs for a range of traits, including yield and yield components (Agrama et al., 2007), cold tolerance (Cui et al., 2013), grain quality, flowering time (Ordonez Jr. et al., 2010) and harvest index (Li et al., 2012). Many associated markers found in these studies were in regions where QTLs have previously been identified. The development of several NGS platforms makes GWAS more and more popular. Using sallow sequencing, a total of 80 MTAs for 14 agronomic traits were identified, explaining an average of around 36% of the phenotypic variance (Huang et al., 2010). Among these loci, six were tied closely to *OsC1* (chromosome 6), *ALK* (chromosome 6), *Rc* (chromosome 6), *qSW5* (chromosome 5) and *GS3* (chromosome 3), respectively. The same method was applied to a larger collection of 950 worldwide rice varieties and identified a total of 32 new loci for flowering time and 10 grain-related traits (Huang et al., 2012). Zhao et al. (2011) performed GWAS using a global collection of 413 rice accessions genotyped using a high-quality custom-designed 44,100 SNP array and phenotyped for 34 traits including morphological, developmental and agronomic traits over two consecutive crop seasons. Dozens of common variants influencing numerous complex traits were identified.

AM is usually conducted using panels of diverse germplasm. While maximizing the genetic diversity, which is beneficial for identifying novel QTL and candidate genes that underlie traits of interest, these panels may not be adapted to identifying relevant variations directly useful in breeding programs (Bordes et al., 2014). Any identified QTL should be validated in a breeding population before they can be used in selection. AM in locally-

adapted breeding material should be of more practical benefit to breeders (Bernardo, 2008), since identified beneficial alleles could be incorporated into the creation of new cultivars with limited deleterious effect (Pauli et al., 2014). Empirical studies in wheat and barley demonstrated that GWAS using elite breeding populations is an effective strategy for integrating new genomic technologies into the development of superior cultivars (Bordes et al., 2014; Mohammadi et al., 2015; Pauli et al., 2014; Pozniak et al., 2012). Recently, GWAS performed using a population of elite *indica* rice breeding lines from IRRI's irrigated breeding program genotyped with GBS identified 52 QTLs for 11 agronomic traits, with some QTLs being co-localized with QTLs identified before (Begum et al., 2015). However, the genome-wide reduction of genetic diversity caused by intensive selection during breeding could reduce the efficiency of AM for some traits. A population mixing lines from several breeding programs, each led by independent breeders with their own germplasm, could increase the level of diversity for an efficient use of AM (Bordes et al., 2014).

The main objective of this study was to identify MTAs/QTLs for GY and 10 related traits using the base breeding population assembled for future *indica* rice breeding for irrigated ecosystem at IRRI. More than 321k markers were generated by the newly developed genotyping-by-sequencing (GBS) method and 76k high quality markers with good genome coverage were used. Population structure of the panel was studied using a model-based Bayesian cluster analysis implemented in STRUCTURE (Falush et al., 2003; Pritchard et al., 2000a), the principal component analysis (PCA) and the nearest neighbour joining (NJ) method (Saitou and Nei, 1987). LD patterns of the whole populations and the inferred subpopulations were investigated. To reduce false positives caused by population structure and unequal relatedness among genotypes (lines) four statistical models were compared using the QQ plot and mean squared difference (MSD) between expected and observed *p* values from analysis proposed by Stich et al. (2008). A total of 452 MTAs that were delineated into 43 QTLs were identified for GY and eight related traits.

5.2 Materials and Methods

Plant materials and phenotyping

Three hundred and ninety two rice varieties or advanced lines developed for the irrigated lowland ecosystem were used in this study to achieve a large amount of genetic diversity. The majority of the lines were from IRRI (225). A good number of lines were also from PhilRice (38), CIAT (19), China (14) and Vietnam (10) (Appendix Table S1). Phenotyping was

conducted in Jiangxi (JX) and Sichuan (SC) in China and IRRI headquarters (Los Baños, Philippines). The experiments in JX and SC were conducted in one crop season in 2012. At IRRI, the experiments were carried out in both dry season (DS) and wet season (WS) of 2012. Three nitrogen fertilizer application rates, no nitrogen, low (90 kg ha^{-1}) and high (180 kg ha^{-1}), were used to create three artificial environments in the DS, designated as DS1, DS2 and DS3, respectively. Similarly, three nitrogen fertilizer application rates, no nitrogen, low (45 kg ha^{-1}) and high (90 kg ha^{-1}), were used to create three artificial environments in the WS, designated as WS1, WS2 and WS3, respectively. GY and 10 related traits including GN, PN, TGW, SN, SR, PB, SB, TN, DTF and PH were measured. The traits measured included GY, grain number per panicle (GN), panicle number (PN), thousand grain weight (TGW), spikelets number per panicle (SN), seed setting rate (SR), number of primary branches per panicle (PB), number of secondary branches per panicle (SB), tiller number per plant (TN), days to flowering (DTF) and plant height (PH). All 11 traits were measured in the experiments conducted at IRRI except for DS2 with only DTF, PH, PN, GY and TGW being measured. GY, DTF, GN, PH, PN, TGW and SR were measured in SC, while GY, DTF, GN, PH and SR were measured in JX. Detailed trial description and trait measurement were given in (Chapter 3).

Genotyping and GBS data analysis

Young leaf tissue was collected from a single plant of 327 lines. DNA extraction was performed using the commercial Qiagen DNeasy 96 Plant kit (cat. no. 69181) following the manufacturer's protocol. Library preparation, sequencing, SNP discovery and calling were performed using 384-plex GBS on the 327 elite inbred lines by the Cornell University using the protocol described by Elshire et al. (2011). Imputation of missing values in the unfiltered SNP data was performed using the software TASSEL 3.0 with default settings. The algorithm works by dividing the SNPs into small SNP windows, then identifying the most similar inbred lines within each window to fill the missing data. The algorithm takes advantage of small identical by decent regions shared between pairs of inbred lines in the collection. If the window from the closest neighbour has more than 5% difference from the line being imputed, the data point is left as missing (Romay et al., 2013). To further improve data quality, other checks were carried out. SNPs with minor allele frequency (MAF) <5% or heterozygosity rate > 5% were discarded after imputation.

Statistical analysis

Population structure

A subset of 1,072 SNPs markers evenly distributed on the 12 chromosomes were chosen to detect population structure of the association panel using a model-based Bayesian clustering analysis method implemented in the software STRUCTURE 2.3.4 (Pritchard et al., 2000a). An admixture model with correlated allele frequencies was used. The number of groups (k) varied from 1 to 10, with burn in 5000 followed by 100,000 Markov Chian Monte Carlo （MCMC）iterations. Fifteen independent runs were conducted for each k. The optimal number of subpopulations was determined based on change of the LnP(D) and the Delta k method proposed by Evanno et al. (2005). PCA was performed with all 76,452 SNPs using GAPIT R package (Lipka et al., 2012). The first three principal components were used to visualize the dispersion of the association panel in a graph. PCA was also analyzed in the subpopultions derived from the STRUCTURE anlaysis. The pairwise similarity coefficients were calculated for all pairwise combinations of the 327 lines according to the method developed by Nei and Li (Nei and Li, 1979) in the software TASSEL v.5.2.11 (Bradbury et al., 2007). The NJ tree based on this gentic similarity matrix was constructed accordingly. To display the familial relationship, a heatmap of the kinship matrix was built up in R (R Core Team, 2015).

Linkage disequilibrium analysis

Genome-wide LD analysis was performed for the whole panel and the two subpopulations. LD was measured using squared allele frequency correlations (r^2) between marker pairs, which was calculated using software TASSEL5.2.11 by setting the sliding window size at 50. The loci were considered to be in significant LD when $p < 0.001$. The pattern and distribution of intra-chromosomal LD within the whole population and each of the two subpopulaiton was visualized and studied from LD plots generated for the whole geome and each of the chromosomes using R (R Core Team, 2015). To investigate the average LD decay in the whole genome significant intra-chromosomal r^2 values were plotted against the distance (bp) between markers. The LD decay was measured as the physical distance (kb) at which r^2 dropped to half of its maximum value (Huang et al., 2012).

Association analysis

Genotypic best linear unbiased estimations (BLUEs) obtained from the single-site phenotypic analysis (Chapter 3) was used as phenotypes for association analysis. Kinship matrix (K)

which measured the genetic similarity between individuals as the proportion of times a given pair of individuals had the same genotype across all SNPs (IBS values) was calculated using TASSEL v.5.2.11 (Bradbury et al., 2007). The following models implemented in TASSEL were tested to choose the best model: i) Q model, a general liner model (GLM) including population structure coefficients derived from STRUCTURE (Q matrix) as covariates; ii) K model, a mixed linear model (MLM) using K to account for unequal genetic relatedness among genotypes; iii) QK model, a MLM including both Q and K correcting for population structure and genetic relatedness; iv) PK model, a MLM including both the first three significant PCs of a PCA of all SNPs as covariates and K. For model comparison, the MSD between the expected and observed *p* values was computed for each model. The model with the smallest MSD was selected as the most appropriate model (Stich and Melchinger, 2009). For each trait, positive false discovery rate (pFDR; q value) was calculated using the QVALUE in R (Storey and Tibshirani, 2003; Storey, 2002). A q value cut off of 0.1 was used for declaring significant associations. Significant markers with r^2 higher than 0.1-0.2 (varying among chromosomes) were delineated into a single QTL.

5.3 Results

Population structure and genetic relatedness between lines

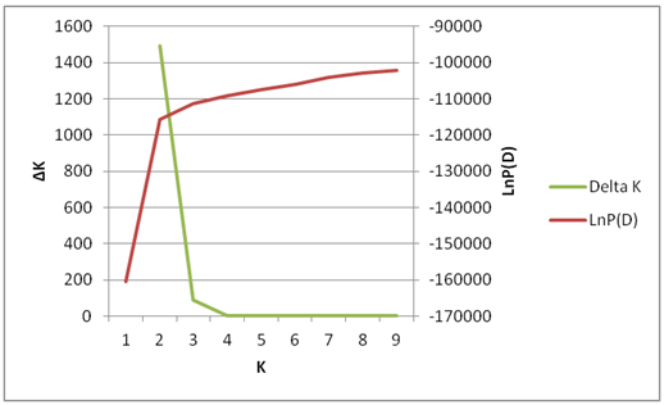
Analysis using a model based approach implemented in STRUCTURE resulted in two optimal subpopulations based on the change of delta k (Figure 5.1a). LnP(D) value was an increasing function of k for all the values observed. However, the biggest increase of LnP(D) was observed when k increased from one to two (Figure 5.1a). Therefore, the association panel was grouped into two subpopulations with 234 and 93 lines, respectively (Figure 5.1b). Table S1 lists the lines in each subpopulation and summarizes the number of lines from different regions/programs of the two subpopulations. All the 93 lines in the subpopulation 2 were from IRRI's breeding programs. All the lines from Pakistan (2), India (3), Africa Rice (4), Bangladesh (6), Indonesia (6), Vietnam (10), China (14), CIAT (19) and PhilRice (38) fell into subpopulation 1 along with 132 IRRI lines.

The PCA analysis showed that the first three PCs explained 15.3%, 3.9% and 2.6% of the total variance, respectively. The panel was grouped into two subpopulations (Figure 5.2a) according to the first PC. The grouping of lines into two subpopulations was the same as the STRUCTURE analysis. PCA analysis conducted within each subpopulation found that the first three PCs explained 5.8% (7.4%), 4.0% (5.7%) and 2.6% (5.2%) of the total variance,

respectively for subpopulation 1 (subpopulation 2) (Figure 5.2b and 5.2c). No obvious substructure was found within either of the subpopulations.

Two distinct subpopulations were also resulted from the NJ tree on the basis of the genetic similarity matrix (Figure 5.3). There was very slight difference in assigning genotypes to subpopulation compared to STRCUTURE and PCA. Four lines in the subpopulation 1 derived from STRUCTURE and PCA analysis were assigned to the subpopulation 2 by the NJ tree.

(a)



(b)

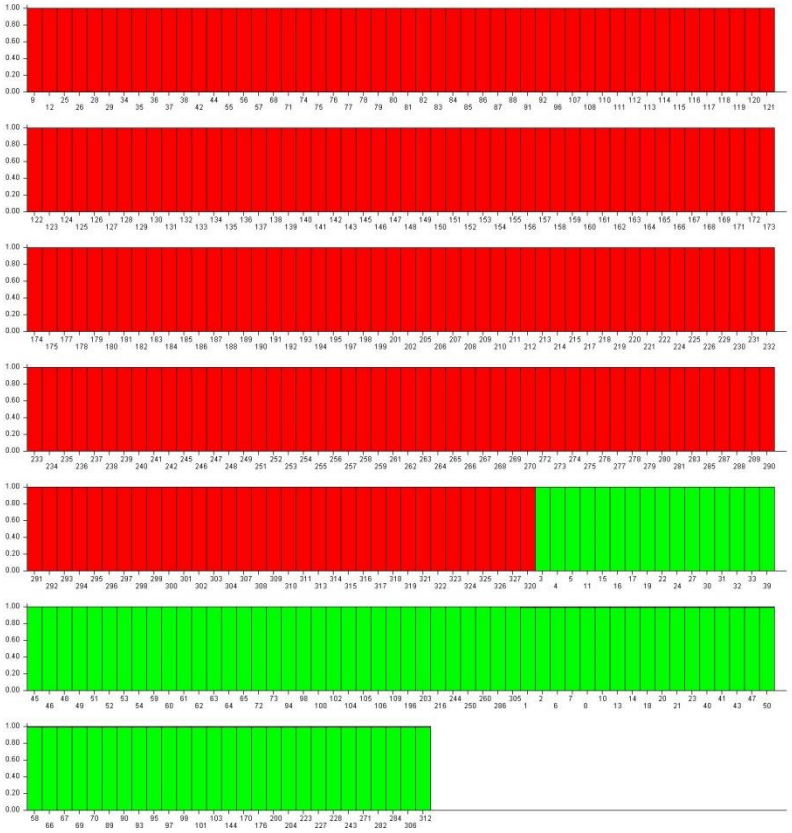


Figure 5.1 Population structure analysis using STRUCTURE based on 1072 SNPs. (a) ΔK and LnP(D) as a function of the number of subpopulations (k). (b) Population sub-structuring.

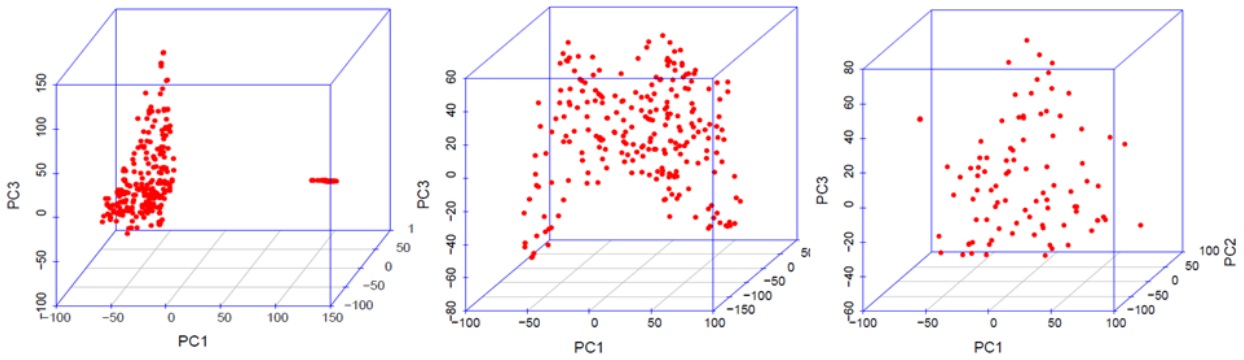


Figure 5.2 Population structure detection using a principal component analysis (PCA). (a) The whole panel of 327 lines, with two populations clearly visible (b) Subpopulation 1 of 234 lines, (c) Subpopulation 2 of 93 lines.

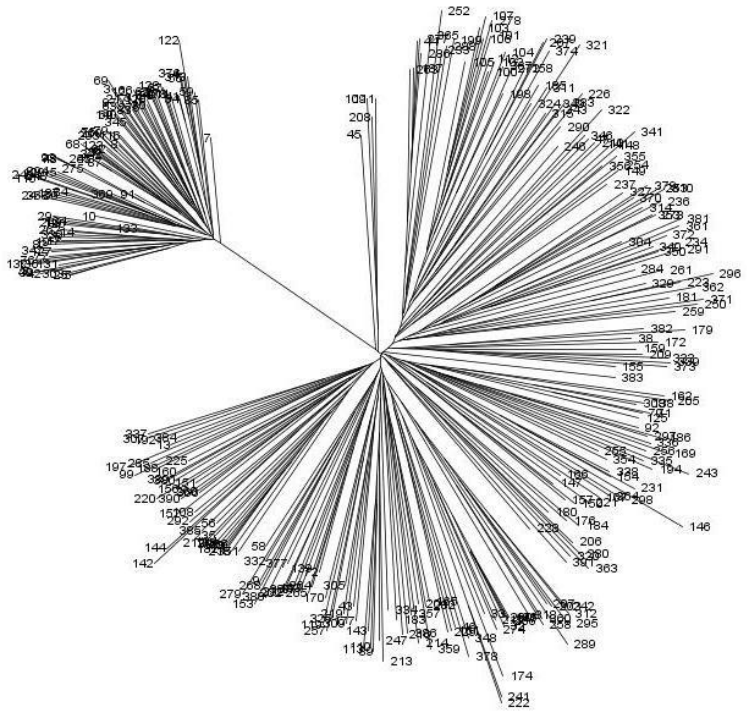


Figure 5.3 Unrooted phylogenetic tree depicting the genetic relations among 327 *indica* lines based on Nei's genetic distance using 76,452 GBS-based SNPs (MAF> 5%).

The kinship coefficients for all the pairwise combinations among the 327 lines ranged from 0 to 2, with a mean of 0.58. There were about 0.18% distantly related genotype pairs with kinship being lower than 0.10 and 0.46% highly related genotype pairs with kinship being higher than 1.50 (Figure 5.4a). The majority of the genotype pairs had kinship similar

to half-sibs (kinship = 0.5). Generally, the relationship within a breeding population was greater than among the breeding populations. The unequal familial relatedness among genotypes could be easily seen in the heatmap (Figure 5.4b).

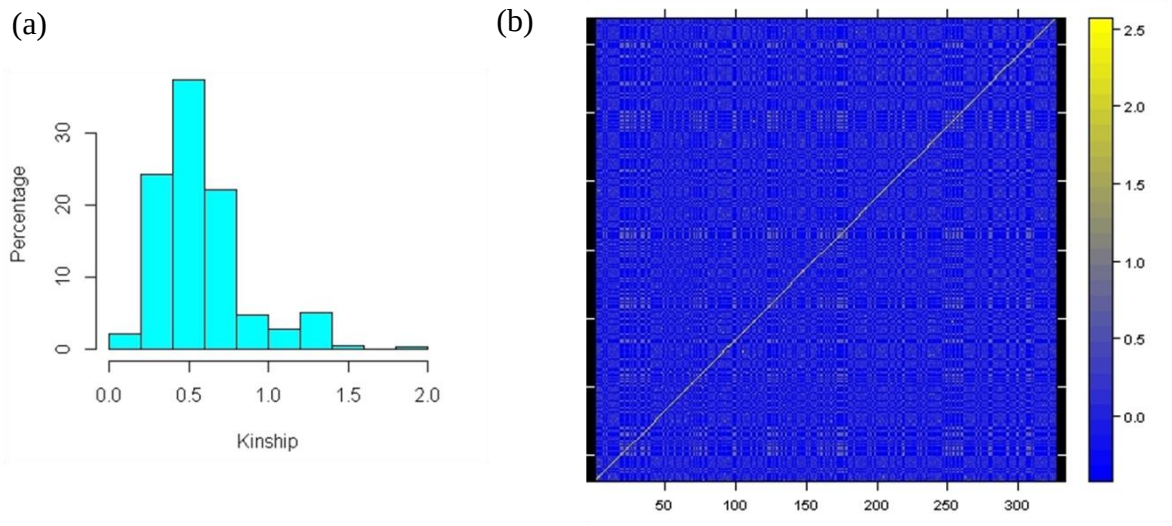


Figure 5.4 Distribution (a) and heatmap (b) of relative kinship coefficient across the 327 lines on 76,452 SNPs.

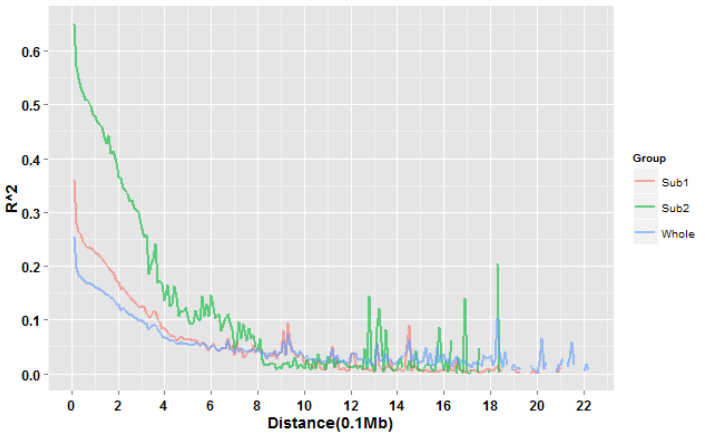
Linkage disequilibrium

The numbers of markers used were 54,188 and 22,261 for subpopulation 1 and subpopulation 2, respectively (Table 5.1). In the whole panel, the r^2 estimate was as high as 0.25 (Figure 5.5a). The mean r^2 among the individual chromosome ranged from 0.21 to 0.34 (Figure 5.5b). LD varied along individual chromosome with regions of high LD interspersing with regions of low LD. As expected, the r^2 value declined with the increasing physical distance between markers. The average r^2 for the whole genome decreased to half its maximum value (0.25) within 200 kb in the whole panel (Figure 5.5a). The LD decay was slower for the chromosomes 8, 12, 1 and 6. The average r^2 dropped to half of its maximum within 250kb for chromosomes 8 and 12 and 210 kb for chromosomes 1 and 6 (Figure 5.5b). LD decay was faster for chromosomes 2, 9 and 10. The average r^2 dropped to half of its maximum within 130 kb for chromosomes 2 and 9 and 150 kb for chromosome 10.

Table 5.1 Summary of the SNP markers analyzed in the whole population and the two subpopulations across all chromosomes (Chr)

Chr	Whole population			Subpopulation 1			Subpopulation 2		
	Size (Mb)	SNP	kb/SNP	Size (Mb)	SNP	Kb/SNP	Size (Mb)	SNP	Kb/SNP
1	43.3	8610	5.0	43.3	5357	8.1	43.3	3252	13.3
2	35.9	8035	4.5	35.9	5342	6.7	35.9	2693	13.3
3	36.4	5041	7.2	36.4	3905	9.3	36.4	1136	32.0
4	35.5	8664	4.1	35.5	6862	5.2	35.5	1802	19.7
5	30.0	5977	5.0	30.0	4199	7.1	30.0	1778	16.8
6	31.1	5629	5.5	31.0	4091	7.6	31.1	1538	20.2
7	29.7	5207	5.7	29.7	3875	7.7	29.6	1331	22.3
8	28.4	7000	4.1	28.4	5094	5.6	28.4	1906	14.9
9	22.9	4908	4.7	22.9	3640	6.3	22.6	1267	17.8
10	23.1	5104	4.5	23.1	3346	6.9	22.8	1758	13.0
11	29.0	6663	4.4	29.0	4666	6.2	29.0	1997	14.5
12	27.5	5614	4.9	27.5	3811	7.2	27.4	1803	15.2
Total	372.8	76452	4.9	372.7	54188	6.9	372.1	22261	16.7

In subpopulation 1 the average r^2 for the whole genome also dropped to half of its maximum (0.36) within 190 kb while it dropped to half of its maximum value (0.65) within 350 Kb in the subpopulation 2 (Figure 5.5a). Similar to the whole panel, the LD decay patterns differed among chromosomes within both of the two subpopulations as well (Figure 5.5b). The maximum r^2 for individual chromosome varied from 0.30 to 0.45 in the subpopulation 1 and from 0.59 to 0.76 in the subpopulation 2 (Figure 5.5b). For subpopulation 1, the LD decay was slower for the chromosomes 12, 3, 4 and 6. The mean r^2 dropped to half of its maximum within 260 kb for chromosome 12, 240 kb for chromosome 4 and 230 kb for chromosomes 4 and 6, respectively (Figure 5.5b). LD decay was faster for chromosomes 9, 10, 11, 2 and 8 with the mean r^2 dropping to half of its maximum value within 100 kb, 130 kb, 130 kb, 140 kb and 140 kb, respectively. For the subpopulation 2, a slower LD decay was observed for chromosomes 1, 8 and 12, with r^2 dropping to half of its maximum value within 490 kb, 420 kb and 400 kb, respectively (Figure 5.5b). For chromosomes 9, 5 and 10, the mean r^2 dropped to half of its maximum within 170 kb, 180 kb and 260 kb, respectively.



(b)

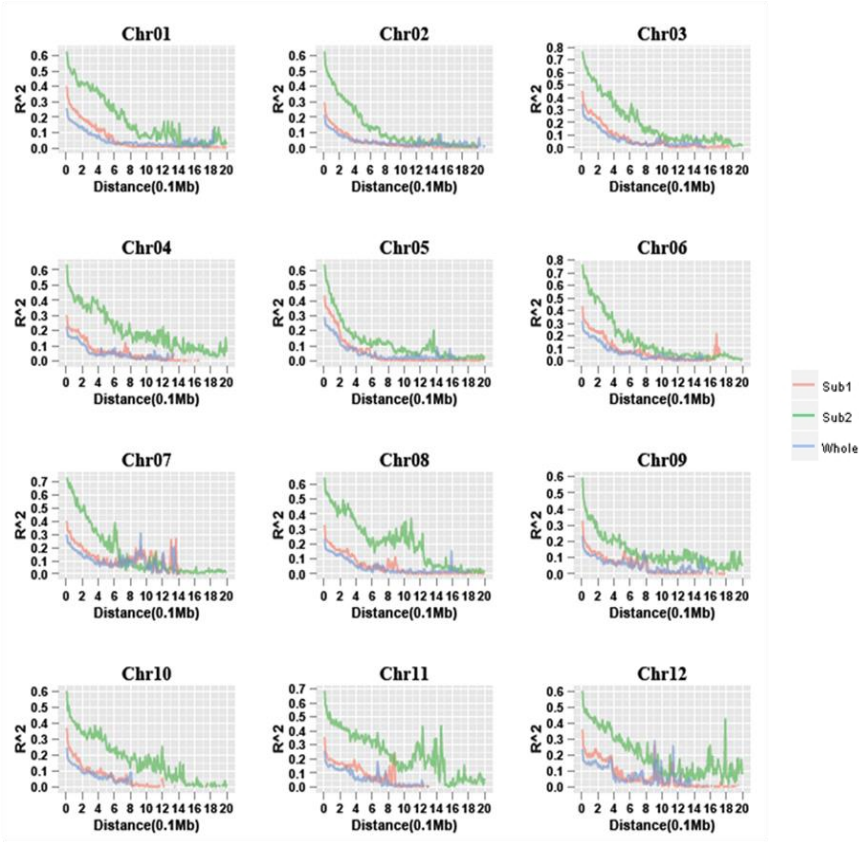


Figure 5.5 LD plot for whole genome (a) and individual chromosome (b) in the whole panel and the two subpopulations inferred by STRUCTURE.

Association analysis

Comparison of models

To choose the most appropriate model for association analysis for all tested traits, four models were tested to detect associations between SNP markers and selected traits from selected environments. Since significant population structure was present in our association panel, only four models considering population structure and/or relatedness among the individuals including Q, K, QK and PK models were tested. The four models were used to detect the associations between SNP markers and DTF in the environments DS2, SC, WS1 and WS2 and DTF, GY, PN and TGW in DS2. Model comparison was made using the QQ plot, which is based on the observed and expected *p* values. The ideal distribution of *p* values should follow a uniform distribution with less deviation from the expected *p* values (Kang et al., 2008). Figure 5.6a shows the QQ plots for the associations between markers and DTF measured in DS2, SC, WS1 and WS2. The *p* values of the Q model significantly deviated from the Y=X line for the four environments. K, QK and PK models showed an improved fit for *p* values and the three models behaved very similarly in the four environments. Figure 5.6b shows the QQ plots for the associations between markers and DTF, GY, PN and TGW measured in DS2. The Q model showed a severe extent of inflation of *p* values for the four traits, especially for PN. For TGW, the *p* values of the Q model were close to the other three models, especially at the high value level. The K, QK and PK models behave very similarly for the four traits in DS2. The result suggested that MLM using either kinship alone or both kinship and population structure could sufficiently account for structure and cryptic familial relatedness among the individuals in the current association panel. Association analysis was then carried out for all the traits in all the testing environments using the K, QK and PK models. MSD indicated that the most appropriate model was the PK model (Stich and Melchinger, 2009). Therefore only the results from the PK model are presented.

Marker trait association

Wide variation was observed for all the tested traits within the whole panel and the two subpopulations based on the average data across the eight testing environments (Figure 5.7 and Table 5.2). Using the PK model, 452 significant MTAs were identified for eight out of 11 tested traits. The Manhattan plots and QQ plots for GY, DTF and PN in DS2 are shown in Figure 5.8 to demonstrate the quality of our GWAS while those for other trait-environment combinations were given in Figure S2. The details of MTAs were given in Table S2. By

delineating markers with r^2 higher than 0.1-0.2 (varying among chromosomes) into a single QTL, there were total of 43 QTLs (Table 5.3-5.5).

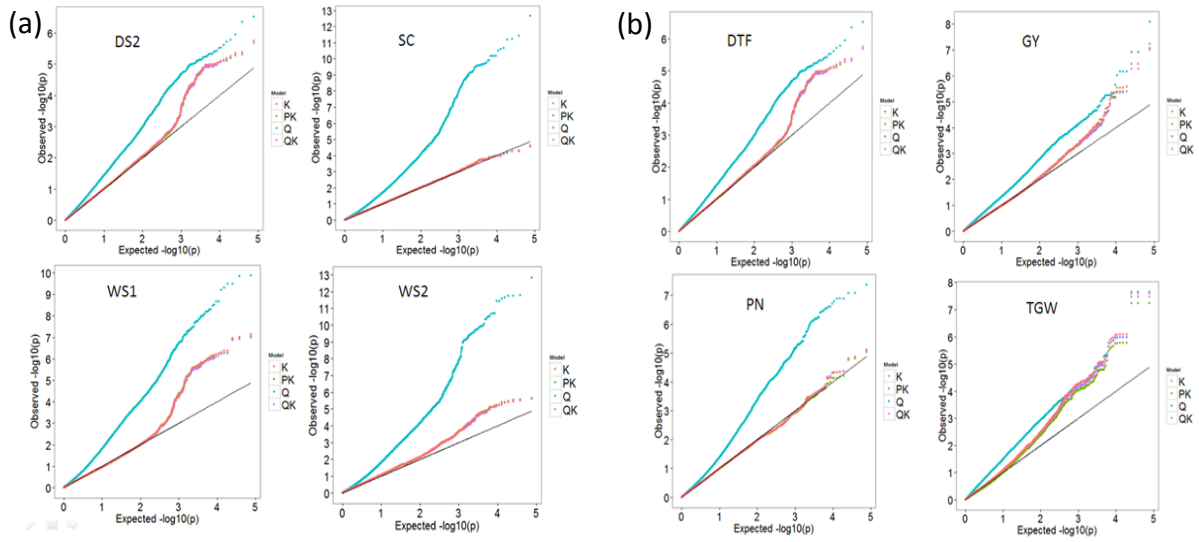


Figure 5.6 Quantile-quantile (QQ) plots of Log p value of four models for GY in environments DS2, SC, WS1 and WS2 (a); Traits DTF, GY, PN and TGW in DS2 (b).
See abbreviations list for environment and trait abbreviations.

Table 5.2 Mean, standard deviation (SD) and range of 11 yield and related traits estimated in the whole panel and the two subpopulations inferred by STRUCTURE

Trait[a]	Whole panel		Subpopulation 1		Subpopulation 2	
	Mean±SD	Range	Mean±SD	Range	Mean±SD	Range
GY	687.6±84.4	429.8-982.9	675.6±83.7	429.8-982.9	717.5±79.0	587.0-962.3
DTF	92±4.9	79-104	91±0.04	79-103	94±4.0	85-104
GN	103.7±13.6	76-147	104.7±14.9	76-147	101.3±9.5	82-125
PB	10.1±0.9	7-13	10.0±0.9	7-12	10.5±0.8	9-13
PH	100.6±7.2	75.2-124.6	99.7±7.6	75.2-124	102.7±5.5	85.8-115.0
PN	9.1±1.7	6-20	8.8±1.6	6-14	9.6±1.9	7-20
SB	24.4±3.6	15-39	24.5±3.8	15-39	24.1±3.0	18-34
SN	124.9±16.9	88-200	125.3±18.7	88-200	23.7±11.4	98-151
SR	82.0±3.8	71.23-91.02	82.4±3.6	72.74-89.20	80.9±34.0	71.23-91.02
TGW	26.69±2.25	18.21-36.43	26.66±2.40	18.21-36.43	26.79±1.8	22.41-32.50
TN	8.7±1.3	5-16	8.6±1.4	5-16	8.8±1.1	7-14

[a] see abbreviations list.

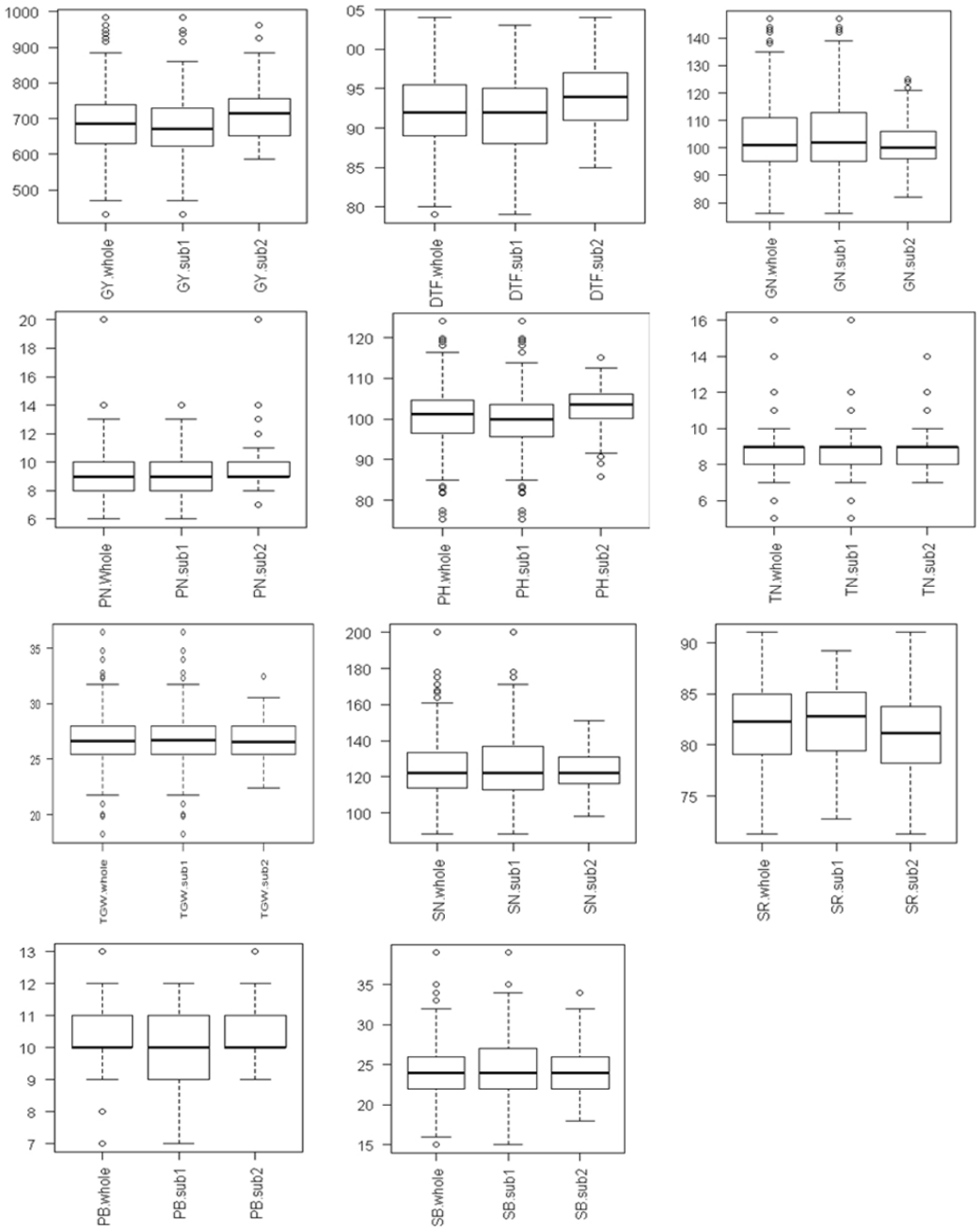


Figure 5.7 Box plots showing the distribution of average BLUEs (y axis) of GY, DTF, PH, GN, PN, PB, SB, SN, SR, TGW and TN within whole panel, subpopulation 1 and subpopulation 2 as defined by STRUCTURE. The trait and population combination indicates under each box.

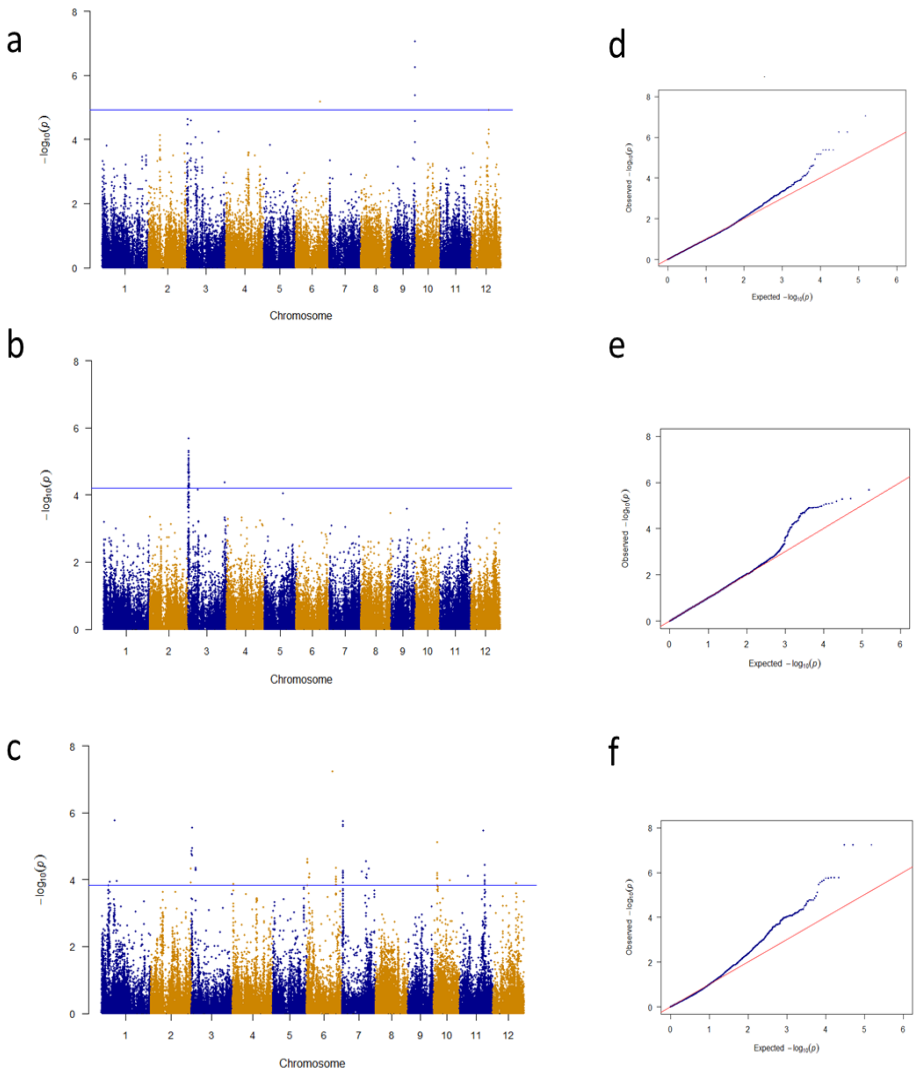


Figure 5.8 Manhattan plots and QQ plots resulting from the GWAS for GY (a, d), DTF (b, e), and PN (c, f) in the environment DS2. Solid lines show the significance threshold for each trait. See abbreviations list for environment and trait descriptors.

Table 5.3 QTLs for GY, GN, PH, SN, TGW, and TN identified by GWAS with q values of less than 0.1

Trait[a]	Env[b]	Marker	Chr[c]	Pos[d]	Interval(nm[e])	QTL	q	R^{2f}	Major[g]	Minor[h]	Effect[i]
GY	DS2	S6_22641374	6	22641374	22.64(3)	*qGY-6*	0.051	0.067	T	C	-295.01
GY	DS2	S9_22230978	9	22230978	22.19-22.23(9)	*qGY-9*	0.014	0.086	G	A	-335.09
GY	DS2	S12_16504535	12	16504535	16.50(1)	*qGY-12*	0.081	0.063	G	A	-320.08
GN	WS1	S2_6491433	2	6491433	6.44-6.56(9)	*qGN-2*	0.016	0.075	T	C	15.87
GN	WS1	S4_33386030	4	33386030	33.35-33.39(2)	*qGN-4*	0.016	0.075	A	G	-29.06
GN	WS1	S8_26716249	8	26716249	26.72(1)	*qGN-8*	0.093	0.058	C	A	-16.24
GN	WS3	S9_22503856	9	22503856	22.50(2)	*qGN-9*	0.045	0.076	A	G	-30.44
PH	DS1	S1_42469942	1	42469942	42.47-42.62(4)	*qPH-1*	0.080	0.078	G	T	4.53
PH	DS1	S9_6666230	9	6666230	6.67-6.71(2)	*qPH-9*	0.087	0.068	C	G	5.44
SN	Ave	S9_19599366	9	19599366	19.60(1)	*qSN-9-1*	0.044	0.070	G	A	-14.27
SN	WS3	S9_22503856	9	22503856	22.50(2)	*qSN-9-2*	0.025	0.080	A	G	-32.66
SN	Ave	S9_22503857	9	22503857	22.50(2)	*qSN-9-2*	0.023	0.079	A	G	-23.97
TGW	SC	S11_20183569	11	20183569	20.00-20.33(22)	*qTGW-11-1*	0.031	0.075	A	G	2.29
TGW	SC	S11_26273587	11	26273587	26.27(1)	*qTGW-11-2*	0.095	0.059	G	T	-2.68
TN	WS3	S2_31613390	2	31613390	30.98-31.75(7)	*qTN-2*	0.013	0.089	G	A	-1.50
TN	WS3	S8_26164899	8	26164899	26.16(1)	*qTN-8*	0.083	0.063	A	T	-1.31
TN	WS2	S9_6918094	9	6918094	6.92(2)	*qTN-9-1*	0.044	0.062	A	T	-1.23
TN	WS2	S9_12223796	9	12223796	12.21-12.27(16)	*qTN-9-2*	0.007	0.087	T	A	-1.77

[a] See abbreviation list for trait descriptors; [b] Env, environment in which the QTL detected.

[c] Chr, chromosome; [d] Pos, physical position (in bp).

[e] nm, number of significant SNP markers within the QTL interval; [f] R^2, variance explained the marker.

[g] Major, major allele; [h] Minor, minor allele.

[i] Effect, effect of the major allele of the most significant marker within the QTL interval.

For GY, three QTLs, *qGY-6*, *qGY-9* and *qGY-12* were identified but only in DS2 (Table 5.3). They were on chromosome 6, 9 and 12 and explained 6.7%, 8.6% and 6.3% of the phenotypic variation, respectively. There were 11 markers within these three QTL regions. The peak SNP for *qGY-6* is a T/C polymorphism with the C allele increasing GY. The peak SNPs for *qGY-9* and *qGY-12* are G/A polymorphisms with the A allele increasing GY. A total of 19, 22 and 16 lines had the increasing alleles, respectively. Four lines, 259, 31, 331 and 330 carry the increasing alleles of all three QTLs. Line 259 ranked first in DS2, and lines 31 and 331 ranked fifth and 23rd, respectively. The yield data for line 330 in DS2 was missing due to rat damage.

Fourteen markers displayed significant associations with GN. These markers were grouped into four QTLs on chromosome 2, 4, 8 and 9, respectively (Table 5.3). *qGN-2*, *qGN-4* and *qGN-8* were identified in WS1 and explained 7.5%, 7.5% and 5.8% of the phenotypic variation, respectively. *qGN-9* was detected in WS3 and accounted for 7.6% of the total phenotypic variation. The peak SNPs for the four QTLs are T/C, A/G, C/A and A/G polymorphisms with the T, G, A and G alleles increasing GN, respectively. There were 264, 23, 33 and 19 lines carrying the GN-increasing alleles of the peak SNPs of the four QTLs, respectively. No line carries the increasing alleles of the peak SNPs of all the four QTLs. Four lines, 75, 124, 225 and 355 that carry the increasing alleles of three of the four QTLs but *qGN-9* ranked 22, 32, 35 and 100 in WS1, respectively.

Two QTL involving six markers were significantly associated with PH (Table 5.3). These two QTLs were on chromosome 1 and 9 and explained 6.8% and 7.8% of the phenotypic variation, respectively. The peak marker of *qPH-1* is a G/T polymorphism with the G allele increasing PH while the marker of *qPH-9* is a C/G polymorphism with the C allele being associated with increased PH.

Three markers on chromosome 9 were found to be associated with SN. These three markers were delineated into two QTLs, *qSN-9-1* and *qSN-9-2* (Table 5.3). The QTL *qSN-9-1* was identified in the average environment and explained 7.0% of the phenotypic variation while *qSN-9-2* was detected in the average environment and WS3, explaining 7.9% and 8.0% of the phenotypic variation, respectively. The peak markers for the two QTLs are both G/A polymorphisms with the A allele of the former and the G allele of latter increasing SN.

Twenty three markers on chromosome 11 were found to be significantly associated with TGW but only in SC (Table 5.3). These markers were delineated into two QTLs explaining 5.9% and 7.5% of the total phenotypic variation, respectively. The peak marker of *qTGW-11-1* is an A/G polymorphism with the A allele increasing TGW while the peak

marker of *qTGW-11-2* is a G/T polymorphism with the T allele increasing TGW. The number of lines carrying increasing allele was 195 and 21, respectively. No line carries both of the increasing alleles.

Twenty-six significantly associated markers in four QTL regions were found for TN (Table 5.3). These QTLs were on chromosome 2, 8, 9 and 9. *qTN-2* and *qTN-8* were identified in WS3, explaining 8.9% and 6.3% of the phenotypic variation, respectively. *qTN-9-1* and *qTN-9-2* were detected in WS2, explaining 6.2% to 8.7% of the phenotypic variation, respectively. The peak marker of *qTN-2* is a G/A polymorphism with the A allele increasing TN. The peak markers of *qTN-8*, *qTN-9-1* and *qTN-9-2* are all A/T polymorphisms with the T allele of the first two and the A allele of the last one increasing TN. The number of lines carrying the increasing allele was 58, 289, 299 and 19, respectively.

For DTF, nine QTLs were identified (Table 5.4). There were 116 markers in the nine QTL regions. Three QTLs were located on chromosome 3. *qDTF-3-1* was associated with DTF in DS2, WS1, WS2 and WS3, explaining 9.4%, 8.1%, 7.2% and 7.1% of the phenotypic variation, respectively. The peak marker of *qDTF-3-1* varied in the four environments. *qDTF-3-2* was associated with DTF in WS1 and explained 4.9% of the phenotypic variation. The peak marker for *qDTF-3-2* is a G/A polymorphism with the A allele being associated with earlier heading. *qDTF-3-3* was detected in DS2 and WS2 and explained 6.0% and 5.3% of the phenotypic variation, respectively. The peak marker for *qDTF-3-3* is a T/C polymorphism with the C allele promoting heading. *qDTF-3-4* was associated with DTF in WS1, WS2 and WS3, explaining 6.6%, 6.2% and 8.5% of the phenotypic variation, respectively. The peak marker of *qDTF-3-4* is a T/A polymorphism with the T allele promoting heading. *qDTF-3-5* was identified in WS1 and WS2, explaining 7.7% and 5.8% of the phenotypic variation, respectively. The peak marker of *qDTF-3-5* is a T/A polymorphism with the T allele promoting heading. *qDTF-3-7* and *qDTF-3-9* were identified in WS1 and explained 5.7% and 5.1% of the phenotypic variation, respectively. The peak markers of *qDTF-3-7* and *qDTF-3-9* are C/T polymorphism with the T allele promoting heading. In JX, one QTL, *qDTF-3-10*, was identified to be associated with DTF and explained 8.3% of the phenotypic variation. The peak SNP is a T/C polymorphism with the C allele promoting heading. Combinations of alleles promoting/delaying heading resulted much earlier/later heading date (HD). For example, for the latest two lines (DTF = 104) in WS1, six of the seven alleles of respective QTLS were associated with late heading.

Table 5.4 QTLs for DTF identified by GWAS with q value less than 0.1

Env[a]	Marker	Chr[b]	Pos[c]	Interval(nm[d])	QTL	q	R^{2e}	Major[f]	Minor[g]	Effect[h]
WS1	S1_22687304	1	22687304	22.69(1)	*qDTF-1*	0.058	0.052	C	G	-3.38
DS2	S3_550877	3	550877	0.26-1.40(204)	*qDTF-3-1*	0.050	0.072	T	G	-5.71
WS3	S3_885458	3	885458	0.26-1.40(204)	*qDTF-3-1*	0.014	0.081	G	T	5.39
WS2	S3_1166378	3	1166378	0.26-1.40(204)	*qDTF-3-1*	0.051	0.071	C	T	-3.72
WS1	S3_1166378	3	1166378	0.26-1.40(204)	*qDTF-3-1*	0.003	0.094	C	T	-5.81
WS1	S3_1682760	3	1682760	1.68(1)	*qDTF-3-2*	0.071	0.049	G	A	3.55
WS2	S3_34390019	3	34390019	34.39(1)	*qDTF-3-3*	0.058	0.060	T	C	4.47
DS2	S3_34390019	3	34390019	34.39(1)	*qDTF-3-3*	0.091	0.053	T	C	4.82
WS3	S4_31516736	4	31516736	31.52(2)	*qDTF-4*	0.013	0.085	T	A	-4.85
WS2	S4_31516736	4	31516736	31.52(2)	*qDTF-4*	0.058	0.062	T	A	-3.80
WS1	S4_31516736	4	31516736	31.52(2)	*qDTF-4*	0.011	0.066	T	A	-4.00
WS2	S5_20558247	5	20558247	20.56(2)	*qDTF-5*	0.062	0.058	T	A	-3.79
WS1	S5_20558247	5	20558247	20.56(2)	*qDTF-5*	0.006	0.077	T	A	-4.81
WS1	S7_20665062	7	20665062	18.77-20.66(2)	*qDTF-7*	0.033	0.057	C	T	3.07
WS1	S9_7844994	9	7844994	7.84(1)	*qDTF-9*	0.052	0.051	C	T	5.54
JX	S10_2734797	10	2734797	2.73-2.93(8)	*qDTF-10*	0.067	0.083	T	C	6.14

[a] Env, environment in which the QTL detected (see abbreviation list)

[b] Chr, chromosome

[c] Pos, physical position (in bp)

[d] nm, number of significant SNP markers within the QTL interval

[e] R^2, variance explained the marker

[f] Major, major allele

[g] Minor, minor allele

[h] Effect, effect of the major allele of the most significant marker within the QTL interval

For PN, 26 QTLs containing 134 markers were identified on chromosome 1, 2, 3, 4, 6, 7, 9, 10 and 11 (Table 5.5). The percentage of phenotypic variation explained ranged from 4.6% (*qPN1-1* in DS2) to 9.6% (*qPN-6-3* in DS2). A single QTL was identified on chromosome 9 (*qPN-9*) and 12 (*qPN-12*), respectively. *qPN-9* and *qPN-12* explained 8.0% and 4.9% of the phenotypic variation, respectively. The peak marker for *qPN-9* is a T/A polymorphism with the A allele increasing PN. The peak SNP for *qPN-12* is an A/C polymorphism with the A allele increasing PN. Two QTLs were identified on chromosome 2 (*qPN-2-1* and *qPN-2-2*), 4 (*qPN-4-1* and *qPN-4-2*) and 10 (*qPN-10-1* and *qPN-10-2*), respectively. *qPN-2-1* was identified in WS2 and WS3, accounting for 5.6% and 8.1% of the phenotypic variation, respectively. Different markers peaked in this QTL region in the two environments. The peak SNP is an A/T polymorphism with the T allele increasing PN in WS2, while is a G/A polymorphism with the A allele increasing PN in WS3. *qPN-2-2* was identified in DS2 and explained 5.3% of the phenotypic variation. The associated SNP is a C/T polymorphism with the T allele increasing PN. *qPN-4-1* was detected in DS2 and explained 4.7% of the phenotypic variation. The peak marker is an A/G polymorphism with the G allele increasing PN. *qPN-4-2* was detected in WS2 and explained 5.8% of the phenotypic variation. The peak marker is a G/T polymorphism with the G allele increasing PN. The two QTLs on chromosome 10 were detected in DS2 and explained 6.4% and 4.8% of the phenotypic variation, respectively. The peak markers are T/C polymorphisms with the C allele increasing PN. Three QTLs were detected on chromosomes 3 (*qPN-3-1*, *qPN-3-2* and *qPN-3-3*) and 11 (*qPN-11-1*, *qPN-11-2* and *qPN-11-3*). *qPN-3-1* and *qPN-3-2* explained 7.0% and 5.3% of the phenotypic variation in DS2, respectively. For *qPN-3-1*, the peak marker is a G/T polymorphism with the T allele increasing PN. For *qPN-3-2*, the peak SNP is a G/C polymorphism with C allele increasing PN. *qPN-3-3* was detected in WS2 and explained 5.9% of the phenotypic variation. The peak marker is an A/G polymorphism with the G allele increasing PN. The three QTLs on chromosome 11 were detected in DS2 and explained 5.0%, 7.0% and 5.4% of the phenotypic variation, respectively. The peak markers are C/T, A/C and G/C polymorphisms with PN increasing allele being T, C and C, respectively. Four QTLs were identified on chromosome 1 (*qPN-1-1*, *qPN-1-2*, *qPN-1-3* and *qPN-1-4*), 6 (*qPN-6-1*, *qPN-6-2*, *qPN-6-3* and *qPN-6-4*) and 7 (*qPN-7-1*, *qPN-7-2*, *qPN-7-3* and *qPN-7-4*), respectively. *qPN-1-1*, *qPN-1-2* and *qPN-1-3* were identified in DS2 and explained 4.7%, 7.4% and 4.8% of the phenotypic variation, respectively. The three corresponding peak markers are all C/T polymorphisms with the T allele increasing PN. *qPN-1-4* was detected in WS2 and explained 7.1% of the phenotypic variation. The peak SNP is a

T/G polymorphism with the G allele increasing PN. The four QTLs on chromosome 6 identified in DS2 explained 5.1% - 9.6% of the phenotypic variation. *qPN-6-3* was also detected in the average environment and explained 8.0% of the phenotypic variation. The corresponding peak markers are all T/C polymorphisms with the C allele increasing PN. The peak marker of *qPN-6-4* is an A/G polymorphism with the G allele increasing PN. Four QTLs on chromosome 7 were identified in DS2 only, and explained 7.3%, 5.2%, 5.6% and 5.5% of the phenotypic variation, respectively. The peak marker, is A/G, A/G, G/C and A/T polymorphisms with the increasing allele being G, A, C and T, respectively.

Multi-trait QTLs were detected in many chromosome regions. Their chromosome positions are shown with other trait-specific QTLs in Figure 5.9. Clusters of QTLs were identified on chromosomes 2, 3, 6, 8, 9, 10 and 11. GY and PN had a common QTL on chromosome 6. GY also shared a QTL on chromosome 9 with GN and SN. Traits DTF and PN shared a common QTL on chromosome 3. DTF had a QTL in common with PN on chromosome 10. Traits PN and TN shared two common QTLs on chromosomes 2 and 9. A QTL region on chromosome 8 was associated with GN and TN. There was a QTL on chromosome 9 for PH and TN.

Table 5.5 QTLs for PN identified by GWAS with q value less than 0.1

Env[a]	Marker	Chr[b]	Pos[c]	Interval(nm[d])	QTL	q	R^{2e}	Major[f]	Minor[g]	Effect[h]
DS2	S1_7023757	1	7023757	6.19-7.02(2)	*qPN-1-1*	0.092	0.047	C	T	-3.13
DS2	S1_11793133	1	11793133	11.79(3)	*qPN-1-2*	0.017	0.074	C	T	-4.31
Ave	S1_11793133	1	11793133	11.79(3)	*qPN-1-2*	0.059	0.066	C	T	-1.77
DS2	S1_13102888	1	13102888	13.10(1)	*qPN-1-3*	0.091	0.048	A	C	-3.76
WS2	S1_18835949	1	18835949	18.84(1)	*qPN-1-4*	0.032	0.071	T	G	-1.59
WS2	S2_31105099	2	31105099	31.11-31.61(2)	*qPN-2-1*	0.097	0.056	A	T	-1.07
WS3	S2_31613390	2	31613390	31.11-31.61(2)	*qPN-2-1*	0.037	0.081	G	A	-1.57
DS2	S2_35210846	2	35210846	35,21-35.24(2)	*qPN-2-2*	0.085	0.053	C	T	-3.28
DS2	S3_519551	3	519551	0.34-0.52(11)	*qPN-3-1*	0.019	0.070	G	T	-4.16
DS2	S3_3849936	3	3849936	3.85(5)	*qPN-3-2*	0.085	0.053	G	C	-3.39
WS2	S3_20032981	3	20032981	20.03(2)	*qPN-3-3*	0.073	0.059	A	G	-1.45
DS2	S4_456956	4	456956	0.46-0.56(4)	*qPN-4-1*	0.095	0.047	A	G	-3.34
WS2	S4_18880870	4	18880870	18.88(1)	*qPN-4-2*	0.091	0.058	G	T	1.21
DS2	S6_658071	6	658071	0.47-0.66(4)	*qPN-6-1*	0.077	0.057	C	T	-3.80
DS2	S6_2174568	6	2174568	2.12-2.19(6)	*qPN-6-2*	0.085	0.051	T	C	-3.16
DS2	S6_22641374	6	22641374	22.64(3)	*qPN-6-3*	0.001	0.096	T	C	-4.90
Ave	S6_22641374	6	22641374	22.64(3)	*qPN-6-3*	0.013	0.080	T	C	-1.80
DS2	S6_26068703	6	26068703	26.07-26.16(15)	*qPN-6-4*	0.085	0.053	A	G	-4.04
DS2	S7_858133	7	858133	0.86-1.16(17)	*qPN-7-1*	0.017	0.073	A	G	-4.39
DS2	S7_17450613	7	17450613	17.45(1)	*qPN-7-2*	0.085	0.052	A	G	4.19
DS2	S7_21459811	7	21459811	21.41-22.14(4)	*qPN-7-3*	0.085	0.056	G	C	-3.89
DS2	S7_23435774	7	23435774	23.44(1)	*qPN-7-4*	0.085	0.055	A	T	-3.65
WS2	S9_12223796	9	12223796	12.21-12.26(16)	*qPN-9*	0.020	0.080	T	A	-1.69
DS2	S10_2888145	10	2888145	2.83-3.0(13)	*qPN-10-1*	0.045	0.064	T	C	-4.23
DS2	S10_14091112	10	14091112	14.09(1)	*qPN-10-2*	0.085	0.048	T	C	-3.07
DS2	S11_7252823	11	7252823	7.25(1)	*qPN-11-1*	0.085	0.050	C	T	-3.45
DS2	S11_20916392	11	20916392	20.92(1)	*qPN-11-2*	0.022	0.070	A	C	-4.62
DS2	S11_22242728	11	22242728	22.01-22.24(15)	*qPN-11-3*	0.085	0.054	G	C	-3.90
DS2	S12_20919811	12	20919811	20.92(1)	*qPN-12*	0.094	0.049	C	A	-3.187

[a] Env, environment in which the QTL detected (see abbreviation list)

[b] Chr, chromosome

[c] Pos, physical position (in bp)

[d] nm, number of significant SNP markers within the QTL interval

[e] R^2, variance explained the marker

[f] Major, major allele

[g] Minor, minor allele

[h] Effect, effect of the major allele of the most significant marker within the QTL interval

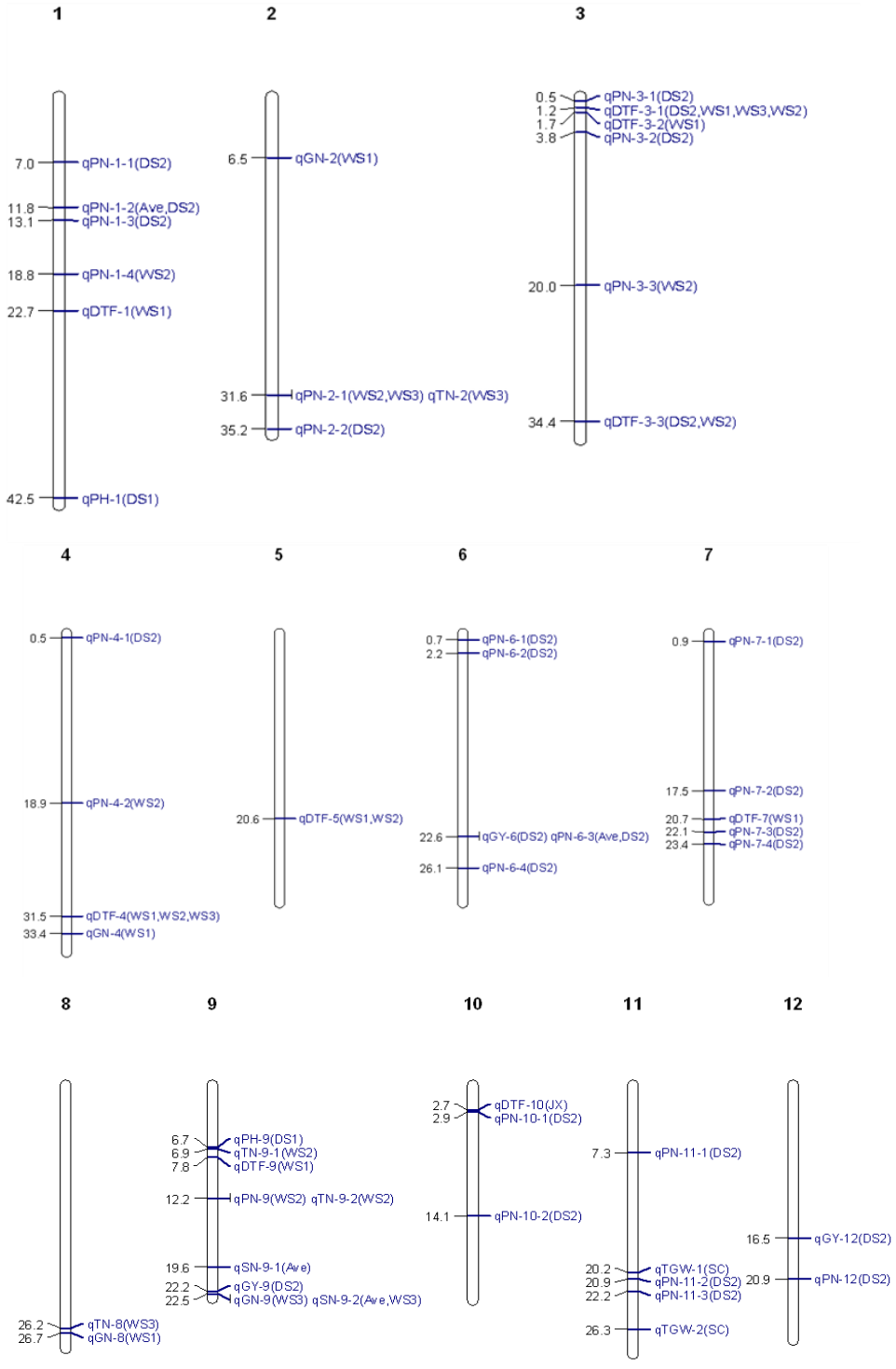


Figure 5.9 Chromosome regions (Mb) of significant QTL identified for traits of interest in this study. The environment in which the QTL identified was given in the bracket; the number on the left of the chromosome indicates the physical position (Mb).

5.4 Discussion

Population structure and unequal familial relationship were present

AM has emerged as a powerful approach for dissecting the genetic basis of complex traits in plants and GWAS is becoming more and more feasible with the advance of high throughput genotyping. However associations detected in AM are often spurious because population structure and unequal relatedness among individuals in a given panel is still remaining a confounding factor (Zhang et al., 2010). In order to reduce false positives, population structure and familial relatedness should be carefully taken into account in AM. The strong population structure and genetic relatedness within plant samples might be affected by many factors such as geographic origins, reproductive nature and artificial selection (Atwell et al., 2010). In the present study, the lines used were from different breeding programs in different parts of the world, with most from IRRI and PhilRice. Although the association panel was from one subspecies, *indica*, the subdivision within the subspecies was still detectable and the population structure still played an important role in QTL identification, as illustrated in the association results from different models with or without population structure coefficients as covariates (see next section). Various numbers of subpopulations were identified from different studies (Wang et al., 2014; Xie et al., 2012; Zhang et al., 2013). With an increased number of markers being used in structure analysis, only two major subpopulations were identified (Wang et al., 2014; Xie et al., 2012). In our study, two consistent subpopulations were detected by all three of methods including STRUCUTRE, PCA and NJ tree. The presence of population structure in populations of *indica* rice varieties has been reported using different panels and marker types (Wang et al., 2014; Xie et al., 2012; Zhang et al., 2013). The existence of population structure within *indica* rice was not out of expectation. In fact, for the majority of crop species and particularly self-pollinated crop species, panels of varieties and advanced breeding lines have population structure and unequal familial relationship (Cockram et al., 2010b; Edae et al., 2014; Flint-Garcia et al., 2005; Pasam et al., 2012), partially caused by the selection for local ecological environments, spatial isolation and assortative mating (i.e. good × good) adopted by breeders.

Big differences were found for the genetic relatedness (estimated as relative kinship) between many pairs of lines (Figure 5.4a and 5.4b). This is expected, since rice breeders usually mate a small number of elite lines in many crosses. Moreover, unequal familial relationship is another important confounding factor in AM using breeding materials.

Linkage disequilibrium was comparable to other panels

Successful GWAS also depends on the extent of LD and the number and distribution of markers. It is recommended that the average marker density should be higher than the LD decay in order to achieve a successful GWAS (Wang et al., 2005; Zhu et al., 2008). Significant LD was detected across the genome of the current rice panel. The patterns of LD varied among chromosomes and subpopulations. For the whole panel the maximum r^2 was 0.25 of the whole genome and ranged from 0.21 to 0.34 among the individual chromosome. Subpopulation 1 showed generally lower LD ($r^2 = 0.36$) than subpopulation 2 ($r^2 = 0.65$) for the whole genome. The genome-wide LD decay in the whole population was around 200 kb, while it was 190 kb in the subpopulation 1 and 350 kb in the subpopulation 2. This was comparable to the estimates reported previously using different *indica* rice panels (Kumar et al., 2015; Mather et al., 2007; McNally et al., 2009; Phung et al., 2014; Xu et al., 2012). In a panel of 220 accessions which consisted mostly of the *indica* accessions, the LD decayed to half of its maximum value at around 300 kb, with the LD ranging from 69 kb on chromosome 3 to 125 kb on chromosome 8 and 12 (Kumar et al., 2015). Significant differences in LD decay were also found in our study, ranging from 130 kb for chromosomes 2 and 9 to 250kb for chromosomes 8 and 12. The differences in LD are mainly caused by the difference in recombination rate, selection history or structural difference among populations/panels. An uneven marker distribution and the difference in marker MAF can also contribute to the large variance associated with estimated LD decay.

Familial relationship was more important than population structure

Several models that were developed to reduce false associations in AM have been successfully applied in previous studies (Price et al., 2006; Pritchard et al., 2000b; Yu et al., 2006). The most commonly used ones in plant association studies are the Q, K, QK and PK models. The Q model accounts for (large-scale) population structure only while the K model takes into account of cryptic familial relatedness only. The QK and PK models were the most stringent because they control both population structure and cryptic familial relatedness. Most of applied plant geneticists have used the QK model, even though the QK model may reduce the power of detecting associations in some cases due to a loss of degrees of freedom (Yu et al., 2006). Indeed, there is not a universal best model for all populations, traits and growing environments.

In the present study, the Q model was significantly worse than the other three models, which behaved similarly. This observation was consistent to Yu et al. (2006) who reported

that the GLM approach with the Q matrix (Q model) was probably more sensitive than the MLM approach (QK) in some cases (Stich et al., 2008a). The degree of the effects of Q matrix and K matrix for controlling false associations was different for different rice panels. Shao et al. (2011) reported that for their whole rice panel including red and black rice, the QK model performed similar to the Q model, with only slightly fewer significant markers identified than the Q model. In the white rice panel, the QK model performed well with 60% reduction in significant markers at different *p* values. Li et al. (2011) found that taking the kinship into account did not improve the mapping QTLs for GY using the USDA rice mini-core collection. The K model performed the worst. The PCA model, which controlled the effect of population structure through PCA, showed the best approximation to the expected cumulative distribution of *p* values, followed by the PK, Q and QK models. This might be due to the fact that the accessions of the mini-core collection were not genealogically related, since the core-collection was developed by maximizing the genetic diversity. In such cases, population structure was the main confounding factor and as a result the models taking into account of population structure performed better than K model which accounts for unequal relatedness only.

For GWAS in rice, we expect that for association panels consisting of breeding lines of a single subspecies/ecotype, unequal relatedness among genotypes is likely to be the main confounding factor and as a result the K model could be sufficient and for panels which consisted of accessions from different subspecies, QK or PK models might be better since both population structure and cryptic relatedness are likely to be important.

Novel QTLs for grain yield and other traits were identified

A total of 43 QTLs for eight out of 11 traits were identified with no QTLs identified for PB, SB and SR. Those QTLs were involved in 18 trait-environment combinations. The numbers of QTLs detected for various traits varied from two (PH, SN and TGW) to 26 (PN). Four of the identified QTLs in the present study were located in the genomic regions where QTLs for yield or related traits have been reported previously. The DTF QTL *qDTF-3-1* detected on chromosome 3 corresponds to the location of *Hd9*, a known QTL for HD (Lin et al., 2002). Another DTF QTL *qDTF-4* identified on chromosome 4 located within the region of previously reported *qGN4-1* for GN (Deshmukh et al., 2010). The PN QTL *qPN-1-1* identified on chromosome 1 was located in the region of known *Gw1-1* for TGW (Yu et al., 2008). The QTL region between 26.2 and 26.7 Mb of chromosome 8, which affected GN and TN, contains the known gene *OsSPL16* (S. Wang et al., 2012), which is synonymous with

GW8 and controls grain size and shape. *OsSPL16* encodes a protein that positively regulates cell proliferation. Over expression of this gene promotes cell division and grain filling, with positive consequences for grain width and GY (S. Wang et al., 2012). Further studies are needed to find whether *OsSPL16* is the underlying candidate gene for GN and TN.

A total of 38 QTLs were found in chromosomal regions where none have been reported previously compared with all mapped QTL, which were registered on the website of Gramene (www.gramene.org). Considering the large number of trait-environment combinations in the present study this number was small. Nevertheless, our results provided further evident that novel QTLs/MTAs could be identified using breeding population. This is the second rice GWAS study using a breeding population as the association panel. Comparing to the first reported study by Begum et al. (2015) using breeding population as association panel, the lines in the present panel are more diverse with two distinct subpopulations. Even though the identified QTLs accounted for a relatively small proportion of the phenotypic variation (R^2 = 4.7% to 9.6%), they should be effective in breeding programs since they were identified from a heterogeneous population with relatively small numbers of lines being from various breeding programs and a stringent statistical model was used.

The current panel of 327 lines is small and has relatively low genetic diversity (Chapter 4), since majority of its lines were from IRRI's breeding program. QTLs with large effects might have been fixed or their MAF were very low. For instances, there were only 22 out of 327 lines that had the favourable allele for *qPN-6-3* which had accounted for the largest proportion of the phenotypic variation (9.6%) among the detected QTLs. For the largest effect QTL for GY on chromosome 9, which explained 8.6% of the phenotypic variation in DS2, only 22 lines carried the favourable allele for increasing GY. In some QTL regions, the LD between the markers and the trait-controlling genes are low due to low marker density. The QTL *qPN-1-1* covered an interval of 0.83Mb on chromosome 1 with 329 markers within this region and the marker density was about 2.5kb/SNP. The QTL *qPN-6-3* only contained three significant markers and the marker density in the 0.83Mb interval is around 8.4kb/SNP. Therefore, it is essential for increasing genetic diversity of the base population (Chapter 4) and for gene discovery to add more diverse lines. One of the options for increasing genetic diversity of the current panel (base breeding population) is to have more lines from large *indica* breeding programs in the subtropical regions such as China (Chapter 4). However, increasing sample size by pooling samples from different breeding programs may not necessarily increase the power to detect MTAs, as demonstrated in wheat

(Mohammadi et al., 2015). MTAs identifiable in one of the population may become undetectable in the combined population. Mohammadi et al. (2015) did identify new QTLs using the combined datasets of three programs. Our own study in rice in a slightly different context also found that many new MTAs were identified by using the whole population compared to the subpopulations, although some MTAs were only identifiable in the subpopulations (Chapter 4).

QTL-by-environment interaction was significant

Testing environments showed significant effects on the association analysis. Among the 43 QTLs identified, 25 QTLs were detected in DS2, 11 in WS2, 10 in WS1, seven in WS3, two in DS1 and SC, respectively, and one in JX. No QTL was detected in DS3. Only three QTLs were detected from the average values from all the environments. The medium nitrogen rate yielded the highest number of significant QTLs, which is 25 in DS2 and 11 in WS2, followed by no nitrogen. The majority of the QTLs for a trait of interest were identified only in one environment with none of them being identified from all the testing environments. QTL-by-environment interaction (QEI) has been reported in previously reported rice GWAS. Zhao et al. (2011) found that only one of the 10 genomic regions associated with candidate genes for flowering time was detected in more than one location and that several GWAS peaks associated with candidate genes were significant in only a single year at the same location. In a different GWAS conducted in two environments using the USDA rice mini-core subset, only six out of the 60 markers detected from two environments were significant in both environments (Nawaz et al., 2015). The environment-specific QTLs could serve as the genetic variability that could be used to produce optimal cultivars once the beneficial stable major genes have been fixed in breeding programs (Pauli et al., 2014).

QTLs affecting multiple traits were present

Eight of the identified QTLs were associated with more than one trait with five were associated with strongly correlated traits measured in the same or different environments. Most of the QTLs for DTF were located on a similar position to these for some of the yield components (PN, GN and TN), suggesting that some maturity genes may have significant effects on yield through pleiotropic effects on yield components. The QTLs for PN on chromosome 6 and QTLs for GN and SN on chromosome 9 were located at the same or similar position to that for GY, thus selecting for these QTLs should be effective in improving GY. They were the QTL region of 22.64 Mb on chromosome 6 (*qGY-6* or *qPN-6*-

3) affecting GY in DS2 and PN in average environment, the genomic region between 31.08 Mb and 31.61 Mb on chromosome 2 (*qPN-2-1* or *qTN-2*) affecting PN in WS2 and WS3 and TN in WS3, the QTL region between 26.16 Mb and 26.71 Mb of chromosome 8 (*qGN-8* or *qTN-8*) affecting GN in WS1 and TN in WS3, QTL region of 12.22 Mb on chromosome 9 (*qPN-9* or *qTN-9-2*) affecting PN and TN in WS2 and the QTL region between 22.23 Mb and 22.50 Mb on chromosome 9 (*qGY-9* or *qGN-9*) affecting GY in DS2 and GN in WS3. Three QTLs affected different traits with a low correlation. They were QTL region between 0.52 Mb and 1.27 Mb on chromosome 3 (*qPN-3-1* or *qDTF-3-1*) affecting PN and DTF in DS2, the genomic region between 6.70 Mb and 6.91 Mb on chromosome 9 (*qPH-9* or *qTN-9-1*) affecting PH in DS1 and TN in WS2, and the genomic region between 2.73 Mb and 2.89 Mb on chromosome 10 (*qDTF-10* or *qPN-10-1*) affecting DTF in JX and PN in DS2. Multi-trait QTLs were often identified in rice genetic studies. Most of the traits sharing same QTLs were closely related. For examples, grain length, width and length-breadth ratio (Begum et al., 2015); surface area of unhusked seeds, length of unhusked seeds and TGW (Yonemaru et al., 2014); filled grains, productive tillers, yield/plant, percent spikelet fertility under stress and percent spikelet fertility SSI (Kumar et al., 2015). Multi-trait QTLs were also found for those showing no obvious relationships, i.e. amylase content and flowering time with blast resistance and flag leaf width (Zhao et al., 2011); HD with surface area of unhusked seeds (Yonemaru et al., 2014). Pleiotropic effects of a single gene and linkage of QTLs for different traits are the biological mechanisms of multi-trait QTLs. Considering that the present panel consisted of lines from different well established breeding program, linkage might be a more likely mechanism. Rice breeders commonly make crosses between similar phenotype (assortative mating) and as a result many linkage blocks were formed and maintained. These linkage blocks also represent untapped genetic variability that could be utilized through breakage of these blocks through recombination, and lead to new genetic gain in populations with unfavourable linkage (Simmonds, 1979). Without cloning or mapping with extremely high resolution it is impossible to tell which of the mechanism is responsible.

Both pleiotropic effects and linkage of the identified QTLs can be either beneficial or troublesome in plant breeding (Guo and Ye, 2014). QTL alleles with favourable effects on different traits are clearly beneficial. Similarly, favourable alleles linked in desirable phase form valuable linkage blocks that would be explored more easily in practical breeding. On the other hand, for QTLs linked in undesirable phase more efforts are needed to break the linkage, which is the main impediment of using QTLs identified from unadapted germplasm in breeding. An allele with favourable effects on some target traits but unfavourable effects

on other traits could be utilized if its negative effects on important traits could be compensated. Attention must be paid to the effects on multiple essential traits of the target QTLs since transferring QTLs even by well-designed MAS is challenging and time consuming (Ye and Smith, 2010).

5.5 Conclusion

A panel of 327 *indica* rice varieties or advanced breeding lines was genotyped by the GBS method and phenotyped for 11 GY related traits in eight environments. The whole panel was divided into two subpopulations. LD decay varied across the chromosomes and the subpopulations. LD decay in the genome was around 200 kb in the whole population. GWAS based on a mixed linear model which incorporates PCA and kinship matrix identified a total of 43 QTLs for eight out of 11 measured traits but PB, SB and SR. A significant interaction existed between QTLs and environments with the majority of the QTLs being only identified in a specific environment. Four QTLs were located in the regions containing known genes/QTLs for yield related traits. Eight multi-trait QTL regions were detected. The effects of identified QTLs were relatively small with the largest percentage of variance explained by a single QTL being 9.6%. The present study demonstrated that MTAs/QTLs could be identified in a heterogeneous population of breeding lines through GWAS even if population structure and unequal familial relationship were present.

Chapter 6 General discussion

6.1 Background

6.1.1 General background

The extreme importance of rice (particularly irrigated rice) to the global economy and food supply for more than half the world's population has been demonstrated in Chapter 1. Therefore increasing rice production is still a big challenge today. Breeding has been a major contributor to rice yield increases in the past (Guimarães, 2009) and enhanced breeding processes will be a key part of that process into the future.

Although rice yield has increased dramatically since the Green Revolution, the increased yield potential of modern rice cultivars has remained stagnant for the past several decades due to the narrowed genetic diversity and various biotic and abiotic stresses) (Caicedo et al., 2007; Nguyen and Ferrero, 2006; Zhu et al., 2007). The commonly used breeding methods, such as pedigree, bulk and backcross, have some disadvantages (Fujimaki, 1980), including limited use of the full range of available genetic resources, restrictions of the potential for genetic recombination, and difficulty of continuing to obtain improvements in successive breeding cycles.

6.1.2 Study scientific environment

To break the rice yield barrier new strategies are being formulated and implemented. This includes but is not limited to targeting varieties to the target population of environments (TPEs) by characterizing genotype-by-environment interaction (GEI), increasing genetic diversity by introducing elite lines from independent breeding programs, wide hybridization to discover and utilize unadapted germplasm, shortening breeding cycles using rapid generation advance methods and implementing MAS or genomic selection (GS). The International Rice Research Institute (IRRI) has proposed an integrated strategy to increase breeding efficiency by utilizing well proven conventional breeding methods, new techniques and methods enabled by modern molecular biology and genomics and advanced methods in experimental design and data analysis (GRiSP, 2010; Ye et al., 2013). This strategy utilizes (marker-assisted) recurrent selection to quickly pyramid the major genes/QTLs in the first few selection cycles and maintain genetic variation contributed by many minor genes to be explored in later cycles, explores GS to reduce the number of breeding cycles and the costs of

phenotyping and adopts advanced experimental design and data analysis methods to improve heritability (Ye et al, 2012). This strategy is being implemented and refined in the "recurrent selection for population improvement program" initiated by IRRI in 2011 (GRiSP, 2010). The studies reported in this thesis fall within this program.

The overall objective of the studies reported by this thesis is to obtain essential information for designing more efficient mating and selection schemes for use in this general breeding strategy.

6.1.3 Study research background

Rice yield is collectively determined by genotype, environments and GEI. The stability of yield performance is one of the most desirable characters of a genotype to be released as a variety, which allows the developed varieties to be adopted across a large area (exploring the general adaption). On the other hand, to achieve maximum productivity requires targeting varieties to their best growing environments (utilizing the specific adaptation). However, GEI for GY of rice grown in irrigated lowland has not received adequate attention comparable to its importance. Data from the "Irrigated-Lowland Rice National Cooperative Testing Program" of the Philippines indicated that that season-by-location interaction (SLI) was highly significant in the combined analyses of variance over seasons and locations, while locations and seasons were not significant (Samonte and Hernandez, 1990 and 1991). Genotype-by-season interaction (GSI) and genotype-by-location interaction (GLI) were significant only in 10 and seven out of the 48 combined analyses, respectively, while genotype by-season-by-location interaction (GSLI) was significant in almost all cases. Datasets from the international irrigated rice yield nursery (IIRYN) conducted in 1993, 1994 and 1995 showed that the GEI sum of squares was 3-7 times the genotype sum of squares variance (INGER, 1993a; 1993b; 1994a; 1994b, 1995a, and 1995b). However, the large GEI could be at least partially caused by the fact that the soil and weather conditions in some of the testing sites were atypical. Therefore, there was only limited information on the magnitude of GEI for GY of rice applicable to a more diverse set of genotypes that has not undergone intense selection for yield stability across diverse environments.

As a model crop species for plant molecular biology and genomics, rice has been sequenced (Eckardt, 2000; IRGSP, 2005) and researchers have accumulated more molecular and genomic information on rice than on most other crops. Thousands of genes/QTLs for grain yield and yield related traits have been identified (Ashikari and Matsuoka, 2006; Huang

et al., 2013; Miura et al., 2011; Xing and Zhang, 2010; Zuo and Li, 2014). Utilizing this genetic information offers the rice breeding community a range of modern tools and methods for addressing the continuing challenge of increasing rice yield. The potential benefits of using molecular markers linked to the genes of interest in breeding programs, which changed from phenotype-based toward a combination of phenotype- and genotype-based selection, have attracted much attention for more than two decades (Bernardo, 2008; Tester and Langridge, 2010). The efficiency and usefulness of marker-assisted selection (MAS) for traits of simple inheritance (i.e. qualitative traits controlled by one or a few genes) have been well proven in many crops, including rice (Collard et al., 2008; Ye and Smith, 2010; Ye et al., 2009). The success of MAS has motivated rice breeders to search for QTLs for complex traits including yield, especially those accounting for a large proportion of phenotypic variation (major QTLs). Many yield-related genes/QTLs have been identified and some of them are fine-mapped or cloned (Xing and Zhang, 2010, Guo and Ye, 2014). The use of these well-characterized genes/QTLs in improving yield has started. However, significant improvement of GY in farm environments has not been reported (Guo and Ye, 2014). For these fine mapped QTLs or cloned genes to make an impact in practical breeding, it is necessary to test their effects in different genetic backgrounds, since the effects of these well characterized genes/QTLs were usually tested using specific populations under specific environments.

The detection of the associations between traits of interest and molecular markers is the prerequisite for MAS. Two main approaches have been used to identify the associations between traits and markers, linkage mapping and association mapping (AM) or linkage disequilibrium (LD) mapping (Darvasi and Shifman, 2005). Although fundamentally different, both approaches share a common strategy that exploits recombination's ability to break up the genome into fragments that can be correlated with phenotypic variation (Myles et al., 2009). Linkage mapping is mainly used to identify those genes segregating in the biparental crosses with contrasting genotypes. AM detects correlations between genotypes and phenotypes in a collection of germplasm based on LD. AM takes advantage of events that created association in a relatively distant past, which has removed association between a QTL and any marker not tightly linked to it due to recombination (Jannick and Walsh, 2002). It has two advantages: broader genetic variation and higher mapping resolution. Therefore, AM is more breeder-friendly and can be used in breeding populations.

6.1.4 Study structure

In this study, a collection of 392 cultivars or advanced *indica* breeding lines were evaluated for GY and 10 other related traits in two target locations in China and two distinct seasons. The cultivars/lines were grown under three different rates of nitrogen fertilizer application at IRRI and the entire data set was used to identify patterns of genotype, environment, and GEI for GY under irrigated ecosystem. Forty-six markers closely linked to 39 cloned or fine-mapped genes/QTLs and 50 random SSR markers which were evenly distributed among the 12 chromosomes were used to genotype the current population. Association analysis was carried out between the 46 markers and the GY and 10 related traits to test the usefulness of the well-known genes/QTLs in this breeding population. The population was also genotyped with genotyping-by-sequencing (GBS) and produced 76K high quality SNPs. Genome-wide association analysis GWAS) was carried out for the 11 tested traits to identify new marker-trait associations (MTAs).

6.2 Overall results

6.2.1 GEI analysis

Ideally, the rice growing environments should be free of severe disease to obtain maximum yield potential of the testing lines. In our experiments, some practices were taken to prevent pest or disease spread, such as, applying Furadan to control golden snail before transplanting and chemical spraying after transplanting. There may have been differences in low level diseases/pest that may contribute to the high GEI. However, we regard them as valid environmental components.

The combined analysis presented in Table 3.1 showed the relatively high magnitude of the GEI variance relative to the genotypic variance for GY. The GEI effect was nearly twice as much as the genotype effect. The GEI for GY was partitioned into principal component axis following the AMMI analysis (Figure 3.1). The first two principal components i.e. IPCA 1 and IPCA 2, which accounted for 68.4% of the total variation, were significant and sufficient to explain the GEI.

The genotype-by-season interaction was found to be the major source of GEI for GY. The DS and WS environments in IRRI were grouped into different groups and discriminated genotypes in different ways. These data (Figure 3.1) demonstrate it is possible to select genotypes with stable performance across seasons. Developing varieties adapted to both of

DS and WS has been the goal of IRRI's irrigated breeding program. However, with distinctive and highly repeatable seasonal pattern and different genotype responses to seasons much larger genetic progress could have been made by breeding for the two seasons separately. Thus, breeding for different seasons separately to exploit the repeatable GEI caused by seasonal changes (as demonstrated in Figure 3.1) was recommended, even with the current breeding gene pools.

It is worth summarizing the similarity of environments in China and at IRRI. Since the rice paddy was irrigated as required in all the testing environments water availability was not a limiting factor for rice growth. The weather conditions including rainfall, temperature and solar radiation were the main source of environmental difference. The maximum temperature of SC and three WS environments was 35 °C. The biplot analysis grouped SC and three WS environments together. Three environments in the DS formed one group with a maximum temperature of 38 °C. JX alone was the third group with a maximum temperature of 37 °C (Figure 3.1b). Thus, IRRI breeding lines with stable and good performance in the WS could be used in SC. Similarly, selection is better to be done in DS in IRRI for use in JX, China. Clearly, great attention should be paid to the relevance of performance at IRRI to their target production environments when IRRI breeding lines are introduced. On the other hand, with a global mandate IRRI's irrigated rice breeding program should expand its testing and selection environments to allow exploiting specific adaption and providing critical and relevant performance information to the developing countries that largely depend on IRRI for new breeding lines.

The different N rates used had only a relatively small effect on the relative performance of genotypes, compared with the season. The contribution of genotype-by-N interaction could become more important if environment determined by location and season has been fixed. Overall, GEI exceeds genotype effects with season and location greater than N fertilizer.

To maximise the value of these results requires that the genes creating the genotype and GEI benefits reported on here need to be identified and tested and that structure and the genetic diversity of the breeding population needs to be assessed. It is necessary to identify new QTLs for GY, which could be used to enlarge the current gene pool. Thus the second and third series of experiments proceeded.

6.2.2 Testing the usefulness of known genes/QTLs for GY and related traits

One of the objectives of the newly designed irrigated breeding program at IRRI is to increase the genetic diversity of its breeding populations by incorporating elite lines from other large breeding programs. The genetic diversity of the base population of 360 lines was studied using all the 46 markers for the 39 targeted genes/QTLs and 53 random SSR markers. It was found (Table 4.2) that the genetic diversity and PIC value of the current population were lower than those of the populations of 83 global *indica* lines (0.694 and 0.665) and 495 Chinese *indica* lines (0.623 and 0.595) assessed by Wang et al. (2014) but higher than the population of 299 inbred *indica* rice varieties mainly from one of the largest *indica* breeding program (Guangdong Academy of Agricultural Sciences) in China (Chapter 4). The base population assembled mainly based on breeder's experiences and phenotypic performance in IRRI still has relatively lower genetic diversity and should be greatly increased by introducing breeding lines from programs with limited germplasm exchange. Based on the following considerations breeding programs in China are particularly interesting. Firstly, the majority of *indica* rice production environment in China is subtropical climates while all other *indica* breeding programs are in tropical regions. Secondly, Chinese *indica* varieties are well-known for high yield. Thirdly, The Green Super Rice program has proven that Chinese varieties performed well in many tropical regions in Asia and Africa, particularly in the dry season.

The usefulness of known genes/QTLs was tested using 39 well characterized genes/QTLs for yield and related traits in the population of *indica* breeding lines for irrigated ecosystem from IRRI and a few other breeding programs. Using association analysis, all the studied genes/QTLs were found to be associated with at least two of the 11 measured traits in one of the eight testing environments or the average environment (Table 4.3). The numbers of genes/QTLs associated with GY, DTF, PH, GN, SN, PN, TN, TGW, PB, SB and SR were 16, 25, 39, 16, 11, five, five, ten, 29, six and 11, respectively. However, all the genes/QTLs were associated with traits unreported previously thus further investigations on the effects of the target genes/QTLs on all important agronomic traits are needed. For all genes/QTLs environment showed large effects and significant gene-by-environment interaction was present. This implies that (1) the effects of the target genes in the TPEs need to be tested before MAS is implemented and (2) It will be difficult to use these genes in breeding for wide adaptation. These genes/QTLs were also associated with two or more traits. It is

important for MAS to investigate the effects of the target genes on other important agronomic traits.

There are some common QTLs in rice, such as *GS3* and *GN1a*. However, the two QTLs were not identified in this study. The fact that those common QTLs couldn't be identified in a GWAS study might be attributed to a few reasons, such as having a fixed QTL, low frequency, low marker density and too small a population. In this study, the minor allele frequencies of the two QTLs were above 5% in the whole population. However, there was no polymorphism of the flanking markers of *GS3* or *GN1a* within each of the subpopulations. So the common QTLs, *GS3* and *GN1a* were not detectable since the model used for association analysis included population structure.

6.2.3 MTAs identification

The main drawback of AM is the high false positive rate caused by population structure and unequal familial relationships between genotypes in the panel. To reduce spurious associations these two factors need to be carefully considered. Using the model based Bayesian cluster method implemented in STRUCTURE, the 360 lines that were selected from 392 lines and genotyped with 53 random SSR markers were grouped into two subpopulations of 205 and 155 lines, respectively (Chapter 4). Three methods were used to detect population structure in the subset of 327 lines genotyped by GBS (Chapter 5). STRUCTURE analysis using 1,072 evenly distributed SNPs on 12 chromosomes grouped these lines into two subpopulations of 234 lines and 93 lines, respectively (Figure 5.1). The same result was obtained by PCA analysis using all the 76452 markers. Separate PCA conducted for the two subpopulations suggested that there was no sub-structure in either of them (Figure 5.2). Near neighbour joining (NJ) method also suggested that there are two subpopulations in the whole panel of 327 lines (Figure 5.3), although four of the lines grouped into the larger subpopulations by STRUCTURE and PCA were assigned to the small population.

The kinship coefficient among the 360 lines was calculated based on the 53 random SSR markers and it ranged from 0 to 1.80, with a mean of 0.49 (Figure 5.4). There were about 2.33% unrelated genotype pairs (kinship = 0), 6.96% distantly related genotype pairs with kinship being lower than 0.10 and 0.03% highly related genotype pairs with kinship being higher than 1.50. For the subset of 327 lines, the kinship coefficients for all the pairwise combinations ranged from 0 to 2, with a mean of 0.58. There were about 0.18%

distantly related genotype pairs with kinship being lower than 0.10 and 0.46% highly related genotype pairs with kinship being higher than 1.50. The majority of the genotype pairs had kinship similar to half-sibs (kinship = 0.5). Generally, the relationship within a breeding population was greater than among the breeding populations.

Successful GWAS also depends on the extent of LD. LD analysis showed that LD decayed to half maximum (r^2=0.25) within a physical distance of 200 kb in the whole population of 327 lines (Figure 5.5). The mean r^2 among the individual chromosome ranged from 0.21 to 0.34. The LD decay was slower for the chromosomes 8, 12, 1 and 6, while it was faster for chromosomes 2, 9 and 10. The average r^2 dropped to half of its maximum within 250kb for chromosomes 8 and 12, and 210 kb for chromosomes 1 and 6. The average r^2 dropped to half of its maximum within 130 kb for chromosomes 2 and 9 and 150 kb for chromosome 10. In subpopulation 1 the average r^2 for the whole genome also dropped to half of its maximum (0.36) within 190 kb while it dropped to half of its maximum value (0.65) within 350 Kb in the subpopulation 2 (Figure 5.5). Similar to the whole panel, the LD decay patterns differed among chromosomes within both of the two subpopulations. The maximum r^2 for individual chromosome varied from 0.30 to 0.45 in the subpopulation 1 and 0.59 to 0.76 in the subpopulation 2. For subpopulation 1, the mean r^2 dropped to half of its maximum within 260 kb for chromosome 12, 240 kb for chromosome 4, and 230 kb for chromosomes 4 and 6, respectively (Figure 5.5). LD decay was faster for chromosomes 9, 10, 11, 2 and 8 with the mean r^2 dropping to its half maximum value within 100 kb, 130 kb, 130 kb, 140 kb and 140 kb, respectively. For the subpopulation 2, a slower LD decay was observed for chromosomes 1, 8, and 12, with r^2 dropping to half its maximum value within 490 kb, 420 kb, and 400 kb, respectively (Figure 5.5). A faster LD decay was observed for chromosomes 9, 5, and 10. The mean r^2 dropped to half its maximum within 170 kb, 180 kb and 260 kb, respectively.

GWAS using a MLM model controlling both population structure and relatedness identified 43 QTLs for all traits except PB, SB, and SR. Three QTLs on chromosome 6, 9 and 12 were identified for GY in DS2 (Table 5.3). The numbers of QTLs identified for the remaining traits varied from two to 26 (Tables 5.4-5.5). Most of the detected QTLs were found in only one environment. Eight of the identified QTLs were associated with more than one trait with five were associated with strongly correlated traits measured in the same or different environments (Figure 5.9). Four of the identified QTLs in the present study were located in the genomic regions where QTLs for yield or related traits have been reported previously. The DTF QTL on chromosome 3, *qDTF-3-1*, identified in four environments

corresponds to *Hd9*, a major QTL for heading date in rice. Some QTLs were located in the regions containing QTLs previously identified for other related traits. One of the PN QTLs on chromosome 1, *qPN-1-1* were in the regions of fine-mapped TGW QTLs, *Gw1-1* and *Gw1-2*. The effects of identified QTLs were relatively small with the highest percentage of phenotypic variance explained by a single QTL being 9.6%.

6.3 Contributions of this study

6.3.1 Overview

Taken as a whole, this study brought together GEI analysis, testing the usefulness of known genes/QTLs for GY and identifying new MTAs for GY and related traits. By having these three elements together it becomes possible to provide useful information for rice breeders to design efficient breeding schemes and incorporate genomic information and tools to accelerate breeding processes.

6.3.2 Summary of outcomes

1. For the first time, a large collection of advanced *indica* lines developed for irrigated ecosystem by many breeding programs were used to study GEI for GY using testing locations in both of tropical and subtropical climates. Artificial environments created by different nitrogen rates were also used, which is one of the important management options for rice production. Genotype, environment, and GEI all significantly affected GY and some of the yield associated traits. GEI was more important than genotypic main effect for GY, SR and PN but less important for other traits. For GY, the genotype-by-season interaction and genotype-by-season-by-nitrogen interaction was more important than the genotype-by-nitrogen interaction. Based on these findings it was suggested that better breeding and testing strategies could be designed to accommodate the effects of large GEI. Such strategies could include;

(1) Adopting multi-environment testing at the early stages of variety development to allow selecting for general adaptation to be conducted earlier. Across-environment performance must be used as selection criterion when sufficient selection intensity can still be applied.

(2) The number of advanced lines tested in different stages of a multi-stage multi-environment testing scheme must be balanced. Testing many lines in the first stages

and a small number of lines at the last stages is not a good option for obtaining reliable GEI information and maximizing genetic gain.

(3) Testing at least 50 lines from a breeding program to include enough genetic variation for the trait of interest.

(4) Subdividing the lowland irrigated ecosystem into more homogeneous TPEs to reduce the effects of GEI if repeatable GEI is identified (Atlin et al., 2000). Breeding separately for the dry and wet seasons in IRRI was recommended.

2. Introduction of IRRI lines has played an important role in many rice breeding programs in developing countries. The importance of choosing lines based on performance in different seasons of IRRI in the introduction of IRRI breeding lines was clearly shown by the two Chinese locations representing two distinct production environments in China. IRRI breeding lines with stable and good performance in the wet season could be used in Sichuan while selection is better done in the dry season in IRRI for use in Jiangxi. Clearly, great attention should be paid to the relevance of performance at IRRI to their target production environments when IRRI breeding lines are introduced. Although the germplasm exchange between China and IRRI has not been extensive for about two decades, there is renewed interest of Chinese and IRRI breeders to utilize the complementary characteristics of IRRI and Chinese breeding gene pools.

3. For the first time, the usefulness of a large number (39) cloned or fine-mapped genes/QTLs for GY and related traits was tested using a population of advanced *indica* breeding lines, which is being used as one of the base populations in IRRI. It was shown that all the 39 target genes/QTLs were associated with at least two tested traits, suggesting that the great efforts of rice geneticists in the last decades provided not only improved understanding of the trait genetics but useful genes for practical breeding as well. The effects of the tested known genes/QTLs were small in the breeding population used, although much larger effects were reported previously using biparental populations, particularly in near isogenic lines (NILs). The prediction of GY using stepwise multiple linear regressions (MLR) with markers for the 39 well characterized genes/QTLs was poor in all eight testing environments. Thus, it is unlikely that significant improvement of GY can be achieved by using one of the well characterized genes/QTLs.

4. It was shown that many of the 39 well characterized genes/QTLs had effects on multiple traits including traits not previously reported and that environment had large effect and significant gene-by-environment interaction was present for all genes/QTLs. Some of the genes/QTLs even had effects in opposite directions in different environments. This is important for using these genes/QTLs in breeding via MAS. Genes must be tested for their effects on all essential agronomic traits in the TPEs before MAS is implemented even if their effects on one or a few traits measured in special environments are clear and functional/diagnostic markers are available. As a model crop species for genomics and molecular biology, more and more genes are being cloned and reported by geneticists and molecular biologists. Breeders must carefully test the effect of such well characterized genes for practical breeding.

5. The base breeding population studied had clear structure with two subpopulations. The results of different statistical methods and different types of markers were very similar. The small subpopulation consisted of lines only from IRRI's breeding program while the large one had lines from other breeding programs and IRRI programs. The presence of population structure in *indica* subspecies is expected and has been reported by other studies (Wang et al., 2014; Xie et al., 2012). However, the IRRI breeding lines were separated into two subpopulations. This was different from a recent report using only breeding lines from the current cycle of IRRI's irrigated breeding program (Begum et al., 2015).

6. The importance of population structure and unequal familial relationships on association analysis was highlighted. It was shown that the unequal familial relationship between genotypes was more important than population structure in the current breeding population, which is consistent with reported association studies using breeding populations in other crops (Bordes et al., 2014).

7. GWAS was applied to the population of advanced *indica* lines and 43 QTLs were identified for GY and seven yield related traits. One QTL for DTF co-localized with *Hd9* for heading date while four QTLs were in regions containing QTLs for other GY related traits. A total of 38 QTLs were novel. Eight multi-trait QTL regions were detected including the region containing the *Hd9*.

6.4 Suggested further studies

1. Characterization of GEI for GY

Significant GEI for GY was observed in this study (Chapter 3). Better breeding and testing strategies to accommodate the effects of large GEI are discussed. However, the study was not aimed at characterizing the TPE of IRRI's irrigated breeding program or the TPE of irrigated lowland ecosystem for *indica* rice in Asia. The number of locations used was few and the trials were conducted only in one year. To characterize the TPE for IRRI's irrigated breeding program requires a representative sample of IRRI's breeding lines to be tested in many more testing locations across multiple years. Subdividing the lowland irrigated ecosystem into more homogeneous TPEs to reduce the effects of GEI will be necessary if repeatable GEI is identified (Atlin et al., 2000). Critical genotypic characters and soil and weather variables that account for a large proportion of GEI will need to be identified to help defining the TPEs.

2. Mining for more favourable alleles of genes/QTLs

Thirty-nine well characterized genes/QTLs for yield and related traits were found to significantly affect two or more yield related traits (Chapter 4). A total of 43 QTLs were identified for GY and seven related traits using high density markers generated by GBS (Chapter 5). However, most of the genes/QTLs only accounted for a small proportion of the total phenotypic variation. It is required that more desirable alleles of these major genes are identified and utilized in breeding. Indeed, a significant portion of the beneficial/superior alleles were left behind during evolution and domestication and as results have not been utilized in modern rice breeding. Introgressions of novel alleles from wild relatives of crop plants into cultivated varieties (DeVicente and Tanksley, 1993; McCouch et al., 2007; J Xiao et al., 1996; Xiao et al., 1998) have clearly demonstrated that certain alleles and their combinations could make dramatic changes in trait expression when moved to a suitable genetic background.

3. Identify new MTAs within the breeding gene pools for irrigated ecosystem

It is necessary to find new genes/QTLs underlying genetics of GY with large effect using a more diverse panel of the breeding gene pools for irrigated ecosystem. There is typically limited germplasm exchange among breeding programs for tropical and subtropical regions. Since different desirable QTLs are present in each of the breeding programs (Pauli et al., 2014), one of the options for increasing genetic diversity of the current panel (base breeding

population) is to have more lines from large *indica* breeding programs in the subtropical regions such as China (Chapter3). Even with similar selection pressure for the same traits, the combination of different parental founders and drift due to small breeding population sizes could lead to differences in the alleles that become dominant in each of the programs. Selection for adaptation to target regions for variety deployment is not similar and could therefore lead to differential selection of yielding alleles through linkage to local adaptation genes, even though selection for improved GY may be similar for all of the programs.

4. Test the effectiveness of MAS using the identified genes/QTLs with relatively large effects

Some of the well characterized genes are useful in our elite breeding population and new QTLs were identified using GWAS in the population. As discussed by Begum et al (2015), the results can be readily used to identify favourable haplotypes that are currently segregating in our population. These haplotypes could then be used to determine the most suitable parents for crossing in order to exploit transgressive segregation and/or to increase the frequency with which favourable haplotypes appear in the progeny. MAS for favourable haplotypes among the progeny would allow increasing breeding efficiency and decreasing cost by reducing the number of plants advanced to the next generation of breeding or that need to be phenotyped.

5. Test the efficiency of GS for GY

The present study showed that the performance of prediction of GY with well-characterized genes/QTLs via MLR method was poor. Therefore it is not likely that those well-characterized genes/QTLs can be used via MAS to improve GY significantly. The usefulness of those well characterized genes/QTLs was tested via association analysis and the results indicated that they were useful in the current *indica* breeding population. However GY couldn't be well predicated with those known genes. With the availability of 76,452 SNPs across the whole genome and phenotypic data in 8 testing environments, it is possible to carry out GS in the current population. The results would provide valuable information for rice breeders.

6. Designing an efficient mating and selection strategy for the general integrated breeding strategy being implemented in IRRI

With so many factors that significantly affect selection response (genetic gain) individually and/or jointly it is not feasible to find the ‘best’ strategies by empirical studies and simulation modeling is necessary (Ye and Van Ginkel, 2011). The information obtained in this study regarding the effects of known genes/QTLs, QEI/GEI, genetic diversity, population structure and LD pattern of the base population and the genomics and molecular biology information gained in the past decades can be used as key inputs in simulation modeling to identify more efficient mating and selection strategy.

References

Abdurakhmonov, I.Y., Abdukarimov, A., 2008. Application of association mapping to understanding the genetic diversity of plant germplasm resources. Int. J. Plant Genomics 574927. doi:10.1155/2008/574927

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Yoshida, K., Mitsuoka, C., Tamiru, M., Innan, H., Cano, L., Kamoun, S., Terauchi, R., 2012. Genome sequencing reveals agronomically important loci in rice using MutMap. Nat. Biotechnol. 30, 174–178. doi:10.1038/nbt.2095

Agarwal, P., Parida, S.K., Mahto, A., Das, S., Mathew, I.E., Malik, N., 2014. Expanding frontiers in plant transcriptomics in aid of functional genomics and molecular breeding. Biotechnol. J. 9, 1480–1492. doi:10.1002/biot.201400063

Agrama, H.A., Eizenga, G.C., Yan, W., 2007. Association mapping of yield and its components in rice cultivars. Mol. Breed. 19, 341–356. doi:10.1007/s11032-006-9066-6

Ahmadi, N., Negrão, S., Katsantonis, D., Frouin, J., Ploux, J., Letourmy, P., Droc, G., Babo, P., Trindade, H., Bruschi, G., Greco, R., Oliveira, M.M., Piffanelli, P., Courtois, B., 2011. Targeted association analysis identified japonica rice varieties achieving Na^+/K^+ homeostasis without the allelic make-up of the salt tolerant indica variety Nona Bokra. Theor. Appl. Genet. 123, 881–95. doi:10.1007/s00122-011-1634-4

Ainley, W.M., Sastry-dent, L., Welter, M.E., Murray, M.G., Zeitler, B., Amora, R., Corbin, D.R., Miles, R.R., Arnold, N.L., Strange, T.L., Simpson, M.A., Cao, Z., Carroll, C., Pawelczak, K.S., Blue, R., West, K., Rowland, L.M., Perkins, D., Samuel, P., Dewes, C.M., Shen, L., Sriram, S., Evans, S.L., Rebar, E.J., Zhang, L., Phillip, D., Urnov, F.D., Webb, S.R., Petolino, J.F., 2013. Trait stacking via targeted genome editing. Plant Biotechnol J. 11, 1126–1134. doi:10.1111/pbi.12107

Alam, R., Sazzadur Rahman, M., Seraj, Z.I., Thomson, M.J., Ismail, A.M., Tumimbang-Raiz, E., Gregorio, G.B., 2011. Investigation of seedling-stage salinity tolerance QTLs using backcross lines derived from Oryza sativa L. Pokkali. Plant Breed. 130, 430–437. doi:10.1111/j.1439-0523.2010.01837.x

Ali, A.J., Xu, J.L., Ismail, A.M., Fu, B.Y., Vijaykumar, C.H.M., Gao, Y.M., Domingo, J., Maghirang, R., Yu, S.B., Gregorio, G., Yanagghihara, S., Cohen, M., Carmen, B., Mackill, D., Li, Z.K., 2006. Hidden diversity for abiotic and biotic stress tolerances in the primary gene pool of rice revealed by a large backcross breeding program. Field

Crops Res. 97, 66–76. doi:10.1016/j.fcr.2005.08.016

Allard, R.W., Bradshaw, A.D., 1964. Implications of genotype-environmental interactions in applied plant breeding. Crop Sci. 4, 503–508.

Aluko, G., Martinez, C., Tohme, J., Castano, C., Bergman, C., Oard, J.H., 2004. QTL mapping of grain quality traits from the interspecific cross Oryza sativa × O. glaberrima. Theor. Appl. Genet. 109, 630–639. doi:10.1007/s00122-004-1668-y

Ando, T., Yamamoto, T., Shimizu, T., Ma, X.F., Shomura, A., Takeuchi, Y., Lin, S.Y., Yano, M., 2008. Genetic dissection and pyramiding of quantitative traits for panicle architecture by using chromosomal segment substitution lines in rice. Theor. Appl. Genet. 116, 881–90. doi:10.1007/s00122-008-0722-6

Arite, T., Iwata, H., Ohshima, K., Maekawa, M., Nakajima, M., Kojima, M., Sakakibara, H., Kyozuka, J., 2007. DWARF10, an RMS1/MAX4/DAD1 ortholog, controls lateral bud outgrowth in rice. Plant J. 51, 1019–29. doi:10.1111/j.1365-313X.2007.03210.x

Ashikari, M., Matsuoka, M., 2006. Identification, isolation and pyramiding of quantitative trait loci for rice breeding. Trends Plant Sci. 11, 344–50. doi:10.1016/j.tplants.2006.05.008

Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., Angeles, E.R., Qian, Q., Kitano, H., Matsuoka, M., 2005. Cytokinin oxidase regulates rice grain production. Science 309, 741–5. doi:10.1126/science.1113373

Asins, M.J., 2002. Present and future of quantitative trait locus analysis in plant breeding. Plant Breed. 121, 281–291. doi:10.1046/j.1439-0523.2002.730285.x

Atlin, G.N., Baker, R.J., McRae, K.B., Lu, X., 2000. Selection response in subdivided target regions. Crop Sci. 40, 7–13. doi:doi:10.2135/cropsci2000.4017

Atwell, S., Huang, Y.S., Vilhjálmsson, B.J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A.M., Hu, T.T., Jiang, R., Muliyati, N.W., Zhang, X., Amer, M.A., Baxter, I., Brachi, B., Chory, J., Dean, C., Debieu, M., de Meaux, J., Ecker, J.R., Faure, N., Kniskern, J.M., Jones, J.D.G., Michael, T., Nemri, A., Roux, F., Salt, D.E., Tang, C., Todesco, M., Traw, M.B., Weigel, D., Marjoram, P., Borevitz, J.O., Bergelson, J., Nordborg, M., 2010. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465, 627–31. doi:10.1038/nature08800

Bai, X., Luo, L., Yan, W., Kovi, M.R., Zhan, W., Xing, Y., 2010. Genetic dissection of rice grain shape using a recombinant inbred line population derived from two contrasting parents and fine mapping a pleiotropic quantitative trait locus qGL7. BMC Genet. 11, 16. doi:10.1186/1471-2156-11-16

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3(10), e3376. doi:10.1371/journal.pone.0003376

Bao, J.S., Corke, H., Sun, M., 2006. Nucleotide diversity in starch synthase IIa and validation of single nucleotide polymorphisms in relation to starch gelatinization temperature and other physicochemical properties in rice (Oryza sativa L.). Theor. Appl. Genet. 113, 1171–1183. doi:10.1007/s00122-006-0355-6

Bao, Q.J., Fan, Y.Y., Yu, W.D., Chen, C., Fan, F.J., Du, J.H., Zhuang, J.Y., 2009. Fine mapping of QTL qPH6-1 for plant height on the short arm of rice chromosome 6. Chinese J. Rice Sci. (Abstract Engligh) 23, 470–474.

Batista, R., Saibo, N., Lourenc, T., Oliveira, M.M., 2008. Microarray analyses reveal that plant mutagenesis may induce more transcriptomic changes than transgene insertion. Proc. Natl. Acad. Sci. U. S. A. 105, 3640–3645. doi:10.1073/pnas.0707881105

Bauchet, M., Mcevoy, B., Pearson, L.N., Quillen, E.E., Sarkisian, T., Hovhannesyan, K., Deka, R., Bradley, D.G., Shriver, M.D., 2007. Measuring European population stratification with microarray genotype data. Am. J. Hum. Genet. 80, 948–956. doi:10.1086/513477

Begum, H., Spindel, J.E., Lalusin, A., Borromeo, T., Gregorio, G., Hernandez, J., Virk, P., Collard, B., McCouch, S.R., 2015. Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (Oryza sativa). PLoS ONE 10(3), e0119873. doi:10.1371/journal.pone.0119873

Bernardo, R., 2008. Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci. 48, 1649. doi:10.2135/cropsci2008.03.0131

Bertero, H.D., de la Vega, A.J., Correa, G., Jacobsen, S.E., Mujica, A., 2004. Genotype and genotype-by-environment interaction effects for grain yield and grain size of quinoa (Chenopodium quinoa Willd.) as revealed by pattern analysis of international multi-environment trials. Field Crops Res. 89, 299–318. doi:10.1016/j.fcr.2004.02.006

Bhasin, H., Bhatia, D., Raghuvanshi, S., Lore, J.S., Sahi, G.K., Kaur, B., Vikal, Y., Singh, K., 2012. New PCR-based sequence-tagged site marker for bacterial blight resistance gene Xa38 of rice. Mol. Breed. 30, 607–611. doi:10.1007/s11032-011-9646-y

Bian, J. M., He, H.H., Li, C.J., Shi, H., Zhu, C.L., Peng, X.S., Fu, J.R., He, X.P., Chen, X.R., Hu, L.F., Ouyang, L.J., 2013. Identification and validation of a new grain weight QTL in rice. Genet. and Molecular Res. 12, 5623-5633.

Bian, J.M., Jiang, L., Liu, L.L., Wei, X.J., Xiao, Y.H., Zhang, L.J., Zhao, Z.G., Zhai, H.Q.,

Wan, J.M., 2010. Construction of a new set of rice chromosome segment substitution lines and identification of grain weight and related traits QTLs. Breed. Sci. 60, 305–313. doi:10.1270/jsbbs.60.305

Biselli, C., Cavalluzzo, D., Perrini, R., Gianinetti, A., Bagnaresi, P., Urso, S., Orasen, G., Desiderio, F., Lupotto, E., Cattivelli, L., Valè, G., 2014. Improvement of marker-based predictability of Apparent Amylose Content in japonica rice through GBSSI allele mining. Rice 7, 1. doi:10.1186/1939-8433-7-1

Blair, M.W., Iriarte, G., Beebe, S., 2006. QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean × wild common bean (Phaseolus vulgaris L.) cross. Theor. Appl. Genet. 112, 1149–63. doi:10.1007/s00122-006-0217-2

Bolle, C., Schneider, A., Leister, D., 2011. Perspectives on systematic analyses of gene function in Arabidopsis thaliana: new tools, topics and trends. Curr. Genomics 12, 1–14. doi:10.2174/138920211794520187

Bonilla, P., Dvorak, J., Mackill, D., Deal, K., Gregorio, G., 2002. RFLP and SSLP mapping of salinity tolerance genes in chromosome 1 of rice (Oryza sativa L.) using recombinant inbred lines. Philipp Agric Sci 85, 68–76.

Borba, T.C. de O., Brondani, R.P.V., Breseghello, F., Coelho, A.S.G., Mendonça, J.A., Rangel, P.H.N., Brondani, C., 2010. Association mapping for yield and grain quality traits in rice (Oryza sativa L.). Genet. Mol. Biol. 33, 515–524. doi:10.1590/S1415-47572010005000065

Bordes, J., Goudemand, E., Duchalais, L., Chevarin, L., Oury, F.X., Heumez, E., Lapierre, A., Perretant, M.R., Rolland, B., Beghin, D., Laurent, V., Le Gouis, J., Storlie, E., Robert, O., Charmet, G., 2014. Genome-wide association mapping of three important traits using bread wheat elite breeding populations. Mol. Breed. 33, 755–768. doi:10.1007/s11032-013-0004-0

Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J.R., Cavalli-Sforza, L.L., 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368, 455–457. doi:10.1038/368455a0

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S., 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23, 2633–5. doi:10.1093/bioinformatics/btm308

Brar, D.S., Khush, G.S., 1997. Alien introgression in rice. Plant Mol. Biol. 35, 35–47.

Breitler, J.C., Vassal, J.M., del Mar Catala, M. Meynard, D., Marfà, V., Melé, E., Royer, M., Murillo, I., San Segundo, B. Guiderdoni, E., Messeguer, J., 2004. Bt rice harbouring cry

genes controlled by a constitutive or wound-inducible promoter, protection and transgene expression under Mediterranean field conditions. Plant Biotechnol J. 2, 417–430.

Breseghello, F., Sorrells, M.E., 2006. Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics 172, 1165–1177. doi:10.1534/genetics.105.044586

Broman, K.W., 2001. Review of statistical methods for QTL mapping in experimental crosses. Lab Anim. (NY). 30, 44–52.

Brown, T.A., Jones, M.K., Powell, W., Allaby, R.G., 2009. The complex origins of domesticated crops in the Fertile Crescent. Trends Ecol. Evol. 24, 103–109. doi:10.1016/j.tree.2008.09.008

Buckler, E.S., Thornsberry, J.M., 2002. Plant molecular diversity and applications to genomics. Curr. Opin. Plant Biol. 5, 107–111. doi:10.1016/S1369-5266(02)00238-8

Caicedo, A.L., Williamson, S.H., Hernandez, R.D., Boyko, A., Fledel-Alon, A., York, T.L., Polato, N.R., Olsen, K.M., Nielsen, R., McCouch, S.R., Bustamante, C.D., Purugganan, M.D., 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genet. 3(9), e163. doi:10.1371/journal.pgen.0030163

Cao, L., Zhan, X., 2014. Chinese experiences in breeding three-line, two-line and super hybrid rice, in: Yan, W., Bao, J. (Eds.), Rice - Germplasm, Genetics and Improvement. InTech, pp. 279–308.

Cao, L.-Y., Wu, J.-L., Fan, Y.-Y., Cheng, S.-H., Zhuang, J.-Y., 2010. QTL analysis for heading date and yield traits using recombinant inbred lines of indica rice grown in different cropping seasons. Plant Breed. 129, 676–682. doi:10.1111/j.1439-0523.2010.01772.x

Carlborg, O., Leif, A., Kinghorn, B., 2000. The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci. Genetics 155, 2003–2010.

Cassman, K.G., Dobermann, A., Walters, D.T., Yang, H., 2003. Meeting cereal demand while protecting natural resources and improving environmental quality. Annu. Rev. Environ. Resour. 28, 315–358. doi:10.1146/annurev.energy.28.040202.122858

Ceccarelli, S., 1996. Adaptation to low/high input cultivation. Euphytica 92, 203–214. doi:10.1007/BF00022846

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N. V, Bogdanove, A.J., Voytas, D.F., 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic

Acids Res. 39, e82. doi:10.1093/nar/gkr218

Chakhonkaen, S., Pitnjam, K., Saisuk, W., Ukoskit, K., Muangprom, A., 2012. Genetic structure of Thai rice and rice accessions obtained from the International Rice Research Institute. Rice 5, 19. doi:10.1186/1939-8433-5-19

Chandler, R.F., 1979. Rice in the Tropics: A Guide to the Development of National Programs. Westview Press, Boulder, Colorado.

Châtel, M., Ospina, Y., Rodriguez, F., Lozano, V.H., Delgado, H., 2008. Upland rice composite population breeding and selection of promising lines for Colombian savannah ecosystem. Pesqui. Agropecuária Trop. 38, 1–5.

Chauhan, H., Khurana, N., Agarwal, P., Khurana, P., 2011. Heat shock factors in rice (Oryza sativa L.): genome-wide expression analysis during reproductive development and abiotic stress. Mol. Genet. Genomics 286, 171–187. doi:10.1007/s00438-011-0638-8

Chen, H., Xie, W., He, H., Yu, H., Chen, W., Li, J., Yu, R., Yao, Y., Zhang, W., He, Y., Tang, X., Zhou, F., Deng, X.W., Zhang, Q., 2014. A high-density SNP genotyping array for rice biology and molecular breeding. Mol. Plant 7, 541–553. doi:10.1093/mp/sst135

Chen, S., Lin, X.H., Xu, C.G., Zhang, Q., 2000. Improvement of bacterial blight resistance of "Minghui 63", an elite restorer line of hybrid rice, by molecular marker-assisted selection. Crop Sci. 40, 239–244. doi:10.2135/cropsci2000.401239x

Chen, Z., Hu, F., Xu, P., Li, J., Deng, X., Zhou, J., Li, F., Chen, S., Tao, D., 2009. QTL analysis for hybrid sterility and plant height in interspecific populations derived from a wild rice relative, Oryza longistaminata. Breed. Sci. 59, 441–445. doi:10.1270/jsbbs.59.441

Cheng, S.-H., Zhuang, J.-Y., Fan, Y.-Y., Du, J.-H., Cao, L.-Y., 2007. Progress in research and development on hybrid rice: A super-domesticate in China. Ann. Bot. 100, 959–966. doi:10.1093/aob/mcm121

Christou, P., Ford, T.L., Kofron, M., 1991. Production of transgenic rice (Oryza sativa L.) plants from agronomically important indica and japonica varieties via electric discharge particle acceleration of exogenous DNA into immature zygotic embryos. Nat. Biotechnol. 9, 957–962. doi:doi:10.1038/nbt1091-957

Clarke, G., Stefanova, K., 2011. Optimal design for early generation plant breeding trials with unreplicated or partially replicated test lines. Aust. N. Z. J. Stat. 53, 461–480. doi:10.1111/j.1467-842X.2011.00642.x

Cockram, J., White, J., Zuluaga, D.L., Smith, D., Comadran, J., Macaulay, M., Luo, Z., Kearsey, M.J., Werner, P., Harrap, D., Tapsell, C., Liu, H., Hedley, P.E., Stein, N.,

Schulte, D., Steuernagel, B., Marshall, D.F., Thomas, W.T.B., Ramsay, L., Mackay, I., Balding, D.J., Waugh, R., O'Sullivan, D.M., 2010a. Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. Proc. Natl. Acad. Sci. U. S. A. 107, 21611–21616. doi:10.1073/pnas.1010179107

Cockram, J., White, J., Zuluaga, D.L., Smith, D., Comadran, J., Macaulay, M., Luo, Z., Kearsey, M.J., Werner, P., Harrap, D., Tapsell, C., Liu, H., Hedley, P.E., Stein, N., Schulte, D., Steuernagel, B., Marshall, D.F., Thomas, W.T.B., Ramsay, L., Mackay, I., Balding, D.J., Waugh, R., O'Sullivan, D.M., 2010b. Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. Proc. Natl. Acad. Sci. U. S. A. 107, 21611–21616. doi:10.1073/pnas.1010179107

Collard, B.C.Y., Mackill, D.J., 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos. Trans. R. Soc. Biol. Sci. 363, 557–72. doi:10.1098/rstb.2007.2170

Collard, B.C.Y., Vera Cruz, C.M., McNally, K.L., Virk, P.S., Mackill, D.J., 2008. Rice molecular breeding laboratories in the genomics era: current status and future considerations. Int. J. Plant Genomics 2008, 524847. doi:10.1155/2008/524847

Comstock, R.E., 1977. Quantitative genetics and the design of breeding programs, in: Proceedings of the International Conference on Quantitative Genetics, August 16-21, 1976. Iowa State University Press, Ames, USA, pp. 705–718.

Comstock, R.E., Moll, R.H., 1963. Genotype × Environment Interactions, in: Symposium on Statistical Genetics and Plant Breeding. National Academy Science National Research Council, Washington, D.C., pp. 164–196.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., 2013. Multiplex genome engineering ssing CRISPR/Cas systems. Science 339, 819–823. doi:10.1126/science

Cooper, M., Brennan, P.S., Sheppard, J.A., 1996. A strategy for yield improvement of wheat which accommodates large genotype by environment interactions, in: Cooper, M., Hammer, G.L. (Eds.), Plant Adaptation and Crop Improvement. CAB International/ICRISAT & IRRI, Wallingford, UK, pp. 487–152.

Cooper, M., Fukai, S., Wade, L.J., 1999. How can breeding contribute to more productive and sustainable rainfed lowland rice systems? Field Crops Res. 64, 199–209. doi:10.1016/S0378-4290(99)00060-X

Cooper, M., Rajatasereekul, S., Immark, S., Fukai, S., Basnayake, J., 1999. Rainfed lowland rice breeding strategies for Northeast Thailand. I. genotypic variation and

genotype×environment interactions for grain yield. Field Crops Res. 64, 131–151. doi:10.1016/S0378-4290(99)00056-8

Cooper, M., Somrith, B., 1997. Implications of genotype-by-environment interactions for yield adaptation of rainfed lowland rice: influence of flowering date on yield variation, in: Fukai, S., Cooper, M., Salisbury, J. (Eds.), Breeding Strategies for Rainfed Lowland Rice in Drought-Prone Environments. ACIAR Proceedings 77. pp. 104–114.

Courtois, B., Frouin, J., Greco, R., Bruschi, G., Droc, G., Hamelin, C., Ruiz, M., Clément, G., Evrard, J.C., Van Coppenole, S., Katsantonis, D., Oliveira, M., Negrão, S., Matos, C., Cavigiolo, S., Lupotto, E., Piffanelli, P., Ahmadi, N., 2012. Genetic diversity and population structure in a European collection of rice. Crop Sci. 52, 1663–1675. doi:10.2135/cropsci2011.11.0588

Courtois, B., Greco, R., Bruschi, G., Frouin, J., Ahmadi, N., Hamelin, C., Ruiz, M., Evrard, J., Katsantonis, D., Oliveira, M., Negrao, S., Cavigiolo, S., Lupotto, E., Piffanelli, P., 2011. Molecular characterization of the European rice collection in view of association mapping. Plant Genet. Resour. Charact. Util. 9, 233–235. doi:10.1017/S147926211100027X

Cubas, P., Lauter, N., Doebley, J., Coen, E., 1999. The TCP domain: A motif found in proteins regulating plant growth and development. Plant J. 18, 215–222. doi:10.1046/j.1365-313X.1999.00444.x

Cuevas-Pérez, F.E., Guimarães, E.P., BerrÍo, L., González, D.I., 1992. Genetic base of irrigated rice in Latin America and the Caribbean, 1971 to 1989. Crop Sci. 32, 1054–1059.

Cui, D., Xu, C.-Y., Tang, C.-F., Yang, C.-G., Yu, T.-Q., A, X.-X., Cao, G.-L., Xu, F.-R., Zhang, J.-G., Han, L.-Z., 2013. Genetic structure and association mapping of cold tolerance in improved japonica rice germplasm at the booting stage. Euphytica 193, 369–382. doi:10.1007/s10681-013-0935-x

D'Halluin, K.D., Vanderstraeten, C., Hulle, J. Van, Rosolowska, J., Brande, I. Van Den, Pennewaert, A., Hont, K.D., Bossut, M., Jantz, D., Ruiter, R., Broadhvest, J., 2013. Targeted molecular trait stacking in cotton through targeted double-strand break induction. Plant Biotechnol J. 11, 933–941. doi:10.1111/pbi.12085

Darvasi, A., Shifman, S., 2005. The beauty of admixture. Nat. Genet. 37, 118–119. doi:10.1038/ng0205-118

Darvasi, A., Weinreb, A., Minke, V., Weller, J.I., Soller, M., 1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map.

Genetics 134.

Delseny, M., Salses, J., Cooke, R., Sallaud, C., Regad, F., Lagoda, P., Guiderdoni, E., Ventelon, M., Brugidou, C., Ghesquière, A., 2001. Rice genomics: Present and future. Plant Physiol. Biochem. 39, 323–334. doi:10.1016/S0981-9428(01)01245-1

Des Marais, D.L., Hernandez, K.M., Juenger, T.E., 2013. Genotype-by-environment interaction and plasticity: exploring genomic responses of plants to the abiotic environment. Annu. Rev. Ecol. Evol. Syst. 44, 5–29. doi:10.1146/annurev-ecolsys-110512-135806

Deshmukh, R., Singh, A., Jain, N., Anand, S., Gacche, R., Singh, A., Gaikwad, K., Sharma, T., Mohapatra, T., Singh, N., 2010. Identification of candidate genes for grain number in rice (Oryza sativa L.). Funct. Integr. Genomics 10, 339–347. doi:10.1007/s10142-010-0167-2

Desta, Z.A., Ortiz, R., 2014. Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci. 19, 592–601. doi:10.1016/j.tplants.2014.05.006

DeVicente, M.C., Tanksley, S.D., 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134, 585–596.

Devlin, B., Roeder, K., 1999. Genomic control for association. Biometrics 55, 997–1004.

Dewan, A., Liu, M., Hartman, S., Zhang, S.S.-M., Liu, D.T.L., Zhao, C., Tam, P.O.S., Chan, W.M., Lam, D.S.C., Snyder, M., Barnstable, C., Pang, C.P., Hoh, J., 2006. HTRA1 promoter polymorphism in wet age-related macular degeneration. Science 314, 989–992. doi:10.1126/science.1133807

Dobermann, A., Witt, C., Dawe, D. (Eds.), 2004. Increasing Productivity of Intensive Rice Systems Through Site-Specific Nutrient Management. Enfield, N.H. (USA) and Los Baños (Philippines): Science Publishers, Inc., and International Rice Research Institute (IRRI).

Eckardt, N.A., 2000. Sequencing the rice genome. Plant Cell 12, 2011–2017. doi:10.1105/tpc.12.11.2011

Edae, E.A., Byrne, P.F., Haley, S.D., Lopes, M.S., Reynolds, M.P., 2014. Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. Theor. Appl. Genet. 127, 791–807. doi:10.1007/s00122-013-2257-8

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6(5), e19379. doi:10.1371/journal.pone.0019379

Endo, M., Mikami, M., Toki, S., 2015. Multigene knockout utilizing off-target mutations of

the CRISPR/Cas9 system in rice. Plant Cell Physiol. 56, 41–47. doi:10.1093/pcp/pcu154

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14, 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

Evans, L.T., 1972. Storage capacity as a limitation on grain yield, in: Rice Breeding. IRRI, Los Banos, pp. 499–514.

Falconer, D.S., Mackey, T.F.C., 1996. Introduction to Quantitative Genetics. Addison-Wesley Longman, Harlow, UK.

Falush, D., Stephens, M., Pritchard, J.K., 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164, 1567–1587. doi:10.1111/j.1471-8286.2007.01758.x

Famoso, A.N., Zhao, K., Clark, R.T., Tung, C.-W., Wright, M.H., Bustamante, C., Kochian, L. V, McCouch, S.R., 2011. Genetic architecture of aluminum tolerance in rice (Oryza sativa) determined through genome-wide association analysis and QTL mapping. PLoS Genet. 7(8), e1002221. doi:10.1371/journal.pgen.1002221

Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., Li, X., Zhang, Q., 2006. GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor. Appl. Genet. 112, 1164–1171. doi:10.1007/s00122-006-0218-1

FAOSTAT, 2014. Rice market monitor. doi:10.1044/leader.PPL.19102014.18

Federer, W., Wolfinger, R., 2003. Augmented row- column design and trend analysis., in: Kang, M. (Ed.), Handbook of Formulas and Software for Plant Geneticists and Breeders. Food Products Press, Binghamton, NY.

Fekih, R., Takagi, H., Tamiru, M., Abe, A., Natsume, S., Yaegashi, H., Sharma, S., Sharma, S., Kanzaki, H., Matsumura, H., Saitoh, H., Mitsuoka, C., Utsushi, H., Uemura, A., Kanzaki, E., Kosugi, S., Yoshida, K., 2013. MutMap+: genetic mapping and mutant identification without crossing in rice. PLoS ONE 8(7), e68529. doi:10.1371/journal.pone.0068529

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.-L., Wei, P., Cao, F., Zhu, S., Zhang, F., Mao, Y., Jian-Kang Zhu, 2013. Efficient genome editing in plants using a CRISPR/Cas system. Cell Res. 23, 1229–1232. doi:10.1038/cr.2013.114

Fiorani, F., Schurr, U., 2013. Future scenarios for plant phenotyping. Annu. Rev. Plant Biol. 64, 267–291. doi:10.1146/annurev-arplant-050312-120137

Fischer, K.S., Fukai, S., Kumar, A., Leung, H., Jongdee, B., 2012. Field phenotyping

strategies and breeding for adaptation of rice to drought. Front. Physiol. 3, 1–21. doi:10.3389/fphys.2012.00282

Flint-Garcia, S.A., Thornsberry, J.M., Buckler, E.S., 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54, 357–74. doi:10.1146/annurev.arplant.54.031902.134907

Flint-Garcia, S.A., Thuillet, A.C., Yu, J., Pressoir, G., Romero, S.M., Mitchell, S.E., Doebley, J., Kresovich, S., Goodman, M.M., Buckler, E.S., 2005. Maize association population: A high-resolution platform for quantitative trait locus dissection. Plant J. 44, 1054–1064. doi:10.1111/j.1365-313X.2005.02591.x

Fraley, C., Raftery, A.E., 2007. Model-based methods of classification: using the mclust software in chemometrics. J. Stat. Softw. 18, 1–13.

Frisch, M., Bohn, M., Melchinger, A.M., 1999. Comparison of selection strategies for marker-assisted backcrossing of a gene. Crop Sci. 39, 1295–1301. doi:10.2135/cropsci1999.3951295x

Fujimaki, H., 1980. Recurrent population improvement rice breeding facilitated with male sterility. Gamma F. Symp. 19, 91–101.

Fujita, D., Santos, R.E., Ebron, L.A., Fukuta, Y., Kobayashi, N., 2011. Characterization of quantitative trait locus for days to heading in near-isogenic lines with genetic background of Indica-type rice variety IR64 (Oryza sativa). Plant Breed. 130, 526–532. doi:10.1111/j.1439-0523.2011.01864.x

Fukuoka, S., Saka, N., Koga, H., Ono, K., Shimizu, T., Ebana, K., Hayashi, N., Takahashi, A., Hirochika, H., Okuno, K., Yano, M., 2009. Loss of function of a proline-containing protein confers durable disease resistance in rice. Science 325, 998–1001. doi:10.1126/science.1175550

Fuller, D.Q., 2007. Contrasting patterns in crop domestication and domestication rates: Recent archaeobotanical insights from the old world. Ann. Bot. 100, 903–924. doi:10.1093/aob/mcm048

Furbank, R., Tester, M., 2011. Phenomics-technologies to relieve the phenotyping bottleneck. Trends Plant Sci 16, 635–644. doi:10.1016/j.tplants.2011.09.005

Gaikwad, K.B., Singh, N., Bhatia, D., Kaur, R., Bains, N.S., Bharaj, T.S., Singh, K., 2014. Yield-enhancing heterotic QTL transferred from wild species to cultivated rice Oryza sativa L. PLoS ONE 9(6), e96939. doi:10.1371/journal.pone.0096939

Gao, X., Starmer, J.D., 2008. AWclust: point-and-click software for non-parametric population structure analysis. BMC Bioinformatics 9, 77. doi:10.1186/1471-2105-9-77

Gao, Z., Qian, Q., Liu, X., Yan, M., Feng, Q., Dong, G., Liu, J., Han, B., 2009. Dwarf 88, a novel putative esterase gene affecting architecture of rice plant. Plant Mol. Biol. 71, 265–276. doi:10.1007/s11103-009-9522-x

Gao, Z.-Y., Zhao, S.-C., He, W.-M., Guo, L.-B., Peng, Y.-L., Wang, J.-J., Guo, X.-S., Zhang, X.-M., Rao, Y.-C., Zhang, C., Dong, G.-J., Zheng, F.-Y., Lu, C.-X., Hu, J., Zhou, Q., Liu, H.-J., Wu, H.-Y., Xu, J., Ni, P.-X., Zeng, D.-L., Liu, D.-H., Tian, P., Gong, L.-H., Ye, C., Zhang, G.-H., Wang, J., Tian, F.-K., Xue, D.-W., Liao, Y., Zhu, L., Chen, M.-S., Li, J.-Y., Cheng, S.-H., Zhang, G.-Y., Wang, J., Qian, Q., 2013. Dissecting yield-associated loci in super hybrid rice by resequencing recombinant inbred lines and improving parental genome sequences. Proc. Natl. Acad. Sci. U. S. A. 110, 14492–14497. doi:10.1073/pnas.1306579110

Gauch, G.H., Zobel, R.W., 1996. AMMI analysis of yield trials, in: Kang, M.S., Gauch, H.G. (Eds.), Genotype by Environment Interaction. CRC Press, Boca Raton, FL., pp. 85–122.

Gauch, H.G., 1988. Model selection and validation for yield trials with interaction. Biometrics 44, 705–715.

Gomez, K.A., Gomez, A.A., 1984. Statistical Procedures for Agricultural Research, 2nd ed. John Wiley & Sons Inc., New York. USA.

Gomez-Roldan, V., Fermas, S., Brewer, P.B., Puech-Pagès, V., Dun, E.A., Pillot, J.-P., Letisse, F., Matusova, R., Danoun, S., Portais, J.-C., Bouwmeester, H., Bécard, G., Beveridge, C.A., Rameau, C., Rochange, S.F., 2008. Strigolactone inhibition of shoot branching. Nature 455, 189–194. doi:10.1038/nature07271

Gong, J., Du, J., Fan, Y., Wu, J., Zhuang, J., 2010. Quantitative Trait Loci for Panicle Size and Grain Yield Detected in Interval RM111-RM19 784 on the Short Arm of Rice Chromosome 6. Agric. Sci. China 9, 1085–1092. doi:10.1016/S1671-2927(09)60194-2

Gregory, P.J., Ingram, J.S., Kobayashi, T., 2000. Rice production and global change. Glob. Environ. Res. 2, 71–77.

Grenier, C., Cao, T.-V., Ospina, Y., Quintero, C., Châtel, M.H., Tohme, J., Courtois, B., Ahmadi, N., 2015. Accuracy of genomic selection in a rice synthetic population developed for recurrent selection breeding. PLoS ONE 10(8), e0136594. doi:10.1371/journal.pone.0136594

GRiSP, 2010. Global Rice Science Partnership (GRiSP) Proposal.

Guimarães, E.P., 2009. Rice breeding, in: Carena, M.J. (Ed.), Cereals. The Banks and the Italian Economy, pp. 99–126. doi:10.1007/978-0-387-72297-9

Guo, L., Ma, L., Jiang, H., Zeng, D., Hu, J., Wu, L., Gao, Z., Zhang, G., Qian, Q., 2009.

Genetic analysis and fine mapping of two genes for grain shape and weight in rice. J. Integr. Plant Biol. 51, 45–51. doi:10.1111/j.1744-7909.2008.00793.x

Guo, L., Ye, G., 2014. Use of major quantitative trait loci to improve grain yield of rice. Rice Sci. 21, 65–82. doi:10.1016/S1672-6308(13)60174-2

Guo, Y., Hong, D., 2010. Novel pleiotropic loci controlling panicle architecture across environments in japonica rice (Oryza sativa L.). J. Genet. Genomics 37, 533–544. doi:10.1016/S1673-8527(09)60073-4

Gupta, P.K., Rustgi, S., Kulwal, P.L., 2005. Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol. Biol. 57, 461–485. doi:10.1007/s11103-005-0257-z

Gurimaraes, E.P., 2005. Population Improvement: A Way to Exploiting the Genetic Resources of Latin America. FAO, Rome, Italy.

Hallauer, A.R., 1985. Compendium of recurrent selection methods and their applications. CRC Crit. Rev. Plant Sci. 3, 1–34.

Hayes, B.J., Cogan, N.O.I., Pembleton, L.W., Goddard, M.E., Wang, J., Spangenberg, G.C., Forster, J.W., 2013. Prospects for genomic selection in forage plant species. Plant Breed. 132, 133–143. doi:10.1111/pbr.12037

He, G., Luo, X., Tian, F., Li, K., Zhu, Z., Su, W., Qian, X., Fu, Y., Wang, X., Sun, C., Yang, J., 2006. Haplotype variation in structure and expression of a gene cluster associated with a quantitative trait locus for improved yield in rice. Genome Res. 16, 618–626. doi:10.1101/gr.4814006.7

Henderson, S.A., Fukai, S., Jongdee, B., Cooper, M., 1996. Comparing simulation and experimental approaches to analysis of genotype by environment interactions for yield in rainfed lowland rice, in: Cooper, M., Hammer, G.L. (Eds.), Plant Adaptation and Crop Improvement. CAB International, in association with IRRI and ICRISAT, Wallingford, pp. 443–464.

Hiei, Y., Ohta, S., Komari, T., Kumashiro, T., 1994. Efficient transformation of rice (Oryza sativa L.) mediated by Agrobacterium and sequence analysis of the boundaries of the T-DNA. Plant J. 6, 271–282. doi:10.1046/j.1365-313X.1994.6020271.x

High, S.M., Cohen, M.B., Shu, Q.Y., Altosaar, I., 2004. Achieving successful deployment of Bt rice. Trends Plant Sci. 9, 286– 292. doi:10.1016/j.tplants.2004.04.002

Hirochika, H., 2010. Insertional mutagenesis with Tos17 for functional analysis of rice genes. Breed. Sci. 60, 486–492. doi:10.1270/jsbbs.60.486

Ho, N.H., Baisakh, N., Oliva, N., Datta, K., Frutos, R., Datta, S.K., 2006. Translational

fusion hybrid Bt genes confer resistance against yellow stem borer in transgenic elite Vietnamese rice (Oryza sativa L,) cultivars. Crop Sci. 46, 781–789. doi:10.2135/cropsci2005.06-0111

Hoeschele, I., VanRaden, P.M., 1993. Bayesian analysis of linkage between genetic markers and quantitative trait loci. I. Prior knowledge. Theor. Appl. Genet. 85, 953–960. doi:10.1007/BF00215034

Huang, M., Wu, Y., Tao, X., Liu, Y., Yang, G., Chen, Z., 2015. Genetic diversity of main inbred indica rice varieties applied in Guangdong province as revealed by molecular marker. Rice Sci. 22, 1–8. doi:10.1016/S1672-6308(14)60275-4

Huang, N., Parco, A., Mew, T., Magpantay, G., McCouch, S., Guiderdoni, E., Xu, J., Subudhi, P., Angeles, E.R., Khush, G.S., 1997. RFLP mapping of isozymes, RAPD and QTLs for grain shape, brown planthopper resistance in a doubled haploid rice population. Mol. Breed. 3, 105–113. doi:10.1023/A:1009683603862

Huang, R., Jiang, L., Zheng, J., Wang, T., Wang, H., Huang, Y., Hong, Z., 2013. Genetic bases of rice grain shape: so many genes, so little known. Trends Plant Sci. 18, 218–226. doi:10.1016/j.tplants.2012.11.001

Huang, X., Qian, Q., Liu, Z., Sun, H., He, S., Luo, D., Xia, G., Chu, C., Li, J., Fu, X., 2009. Natural variation at the DEP1 locus enhances grain yield in rice. Nat. Genet. 41, 494–497. doi:10.1038/ng.352

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q., Liu, K., Huang, T., Zhou, T., Jing, Y., Li, W., Lin, Z., Buckler, E.S., Qian, Q., Zhang, Q.-F., Li, J., Han, B., 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. 42, 961–967. doi:10.1038/ng.695

Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., Li, W., Guo, Y., Deng, L., Zhu, C., Fan, D., Lu, Y., Weng, Q., Liu, K., Zhou, T., Jing, Y., Si, L., Dong, G., Huang, T., Lu, T., Feng, Q., Qian, Q., Li, J., Han, B., 2012. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat. Genet. 44, 32–39. doi:10.1038/ng.1018

Ikeda, R., Khush, G.S., Tabien, R.E., 1990. A new resistance gene to bacterial blight derived from O. longistaminata. Japanese J. Breed. 40, 280–281.

INGER, 1993a. Final Report of International Irrigated Yield Nursery (IIRYN-Early). IRRI, Philippines.

INGER, 1993b. Final Report of International Irrigated Yield Nursery (IIRYN-Medium). IRRI,

Philippines.

INGER, 1994a. Final Report of International Irrigated Yield Nursery (IIRYN-Early). IRRI, Philippines.

INGER, 1994b. Final Report of International Irrigated Yield Nursery (IIRYN-Medium). IRRI, Philippines.

INGER, 1995a. Final Report of International Irrigated Yield Nursery (IIRYN-Early). IRRI, Philippines.

INGER, 1995b. Final Report of International Irrigated Yield Nursery (IIRYN-Medium). IRRI, Philippines.

Inthapanya, P., Sipaseuth, Sihavong, P., Sihathep, V., Chanphengsay, M., Fukai, S., Basnayake, J., 2000. Genotype differences in nutrient uptake and utilisation for grain yield production of rainfed lowland rice under fertilised and non-fertilised conditions. Field Crops Res. 65, 57–68. doi:10.1016/S0378-4290(99)00070-2

IRGSP, 2005. The map-based sequence of the rice genome. Nature 436, 793–800. doi:10.1038/nature03895

IRRI, 1996. Standard evaluation system for rice. 4th Edition, IRRI, The Philippines.

IRRI, 2006. Bringing Hope, Improving Lives: Strategic Plan 2007–2015. Los Baños: IRRI.

IRRI, 2013. Standard Evaluation System (SES) for Rice, 5th ed. IRRI, Philippines.

Ishimaru, K., 2003. Identification of a locus increasing rice yield and physiological analysis of its function. Plant Physiolgoy 133, 1083–1090. doi:10.1104/pp.103.027607.

Ishimaru, K., Hirotsu, N., Madoka, Y., Murakami, N., Hara, N., Onodera, H., Kashiwagi, T., Ujiie, K., Shimizu, B.-I., Onishi, A., Miyagawa, H., Katoh, E., 2013. Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. Nat. Genet. 45, 707–711. doi:10.1038/ng.2612

Jackson, M.T., 1997. Conservation of rice genetic resources: the role of the International Rice Genebank at IRRI. Plant Mol. Biol. 35, 61–67.

Jannick, J.L., Walsh, B., 2002. Association mapping in plant populations, in: Kang, M.S. (Ed.), Quantitative Genetics, Genomics and Plant Breeding. CABI, Wallingford.

Jannink, J.-L., 2010. Dynamics of long-term genomic selection. Genet. Sel. Evol. 42, 35. doi:10.1186/1297-9686-42-35

Jannink, J.L., Bink, M.C.A.M., Jansen, R.C., 2001. Using complex plant pedigrees to map valuable genes. Trends Plant Sci. 6, 337–342. doi:10.1016/S1360-1385(01)02017-9

Jannink, J.-L., Walsh, B., 2002. Association mapping in plant populations, in: Kang, M.S. (Ed.), Quantitative Genetics, Genomics and Plant Breeding. CAB International, New

York, pp. 59–68.

Jannink, J.-L., Walsh, B., 2002. Association mapping in plant populations, in: Quantitative Genetics, Genomics and Plant Breeding. CAB International, pp. 59–68.

Jia, L., Yan, W., Zhu, C., Agrama, H.A., Jackson, A., Yeater, K., Li, X., Huang, B., Hu, B., McClung, A., Wu, D., 2012. Allelic analysis of sheath blight resistance with association mapping in rice. PLoS ONE 7, e32703. doi:10.1371/journal.pone.0032703

Jiang, L., Guo, L., Jiang, H., Zeng, D., Hu, J., Wu, L., Liu, J., Gao, Z., Qian, Q., 2008. Genetic analysis and fine-mapping of a dwarfing with withered leaf-tip mutant in rice. J. Genet. Genomics 35, 715–721. doi:10.1016/S1673-8527(08)60226-X

Jiang, S., Zhang, X., Wang, J., Chen, W., Xu, Z., 2010. Fine mapping of the quantitative trait locus qFLL9 controlling flag leaf length in rice. Euphytica 176, 341–347. doi:10.1007/s10681-010-0209-9

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., Weeks, D.P., 2013. Demonstration of CRISPR / Cas9 / sgRNA-mediated targeted gene modification in Arabidopsis , tobacco , sorghum and rice. Nucleic Acids Res. 41, e188. doi:10.1093/nar/gkt780

Jiao, Y., Wang, Y., Xue, D., Wang, J., Yan, M., Liu, G., Dong, G., Zeng, D., Lu, Z., Zhu, X., Qian, Q., Li, J., 2010. Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. Nat. Genet. 42, 541–544. doi:10.1038/ng.591

Jin, F., Kim, D., Ju, H., Ahn, S., 2009. Mapping quantitative trait loci for awnness and yield component traits in isogenic lines derived from an Oryza sativa/O. rufipogon cross. J. Crop Sci. Biotechnol. 10, 9–16. doi:10.1007/s12892-009-0061-4

Jin, L., Lu, Y., Xiao, P., Sun, M., Corke, H., Bao, J., 2010. Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. Theor. Appl. Genet. 121, 475–487. doi:10.1007/s00122-010-1324-7

Juliano, B.O., 1993. Rice in Human Nutrition. Food and Agriculture Organization of the United Nations Rome (Italy) and International Rice Research Institute, Los Baños, Laguna (Philippines). doi:10.1017/CBO9781107415324.004

Kajala, K., Covshoff, S., Karki, S., Woodfield, H., Tolley, B.J., Dionora, M.J.A., Mogul, R.T., Mabilangan, A.E., Danila, F.R., Hibberd, J.M., Quick, W.P., 2011. Strategies for engineering a two-celled C 4 photosynthetic pathway into rice. J. Exp. Bot. 62, 3001–3010. doi:10.1093/jxb/err022

Kang, H.M., Zaitlen, N.A., Wade, C.M., Kirby, A., Heckerman, D., Daly, M.J., Eskin, E., 2008. Efficient control of population structure in model organism association mapping. Genetics 178, 1709–1723. doi:10.1534/genetics.107.080101

Kang, M.S., 1998. Using genotype-environment interaction for crop cultivar development. Adv. Agron. 62, 199–252.

Kang, M.S., 2004. Breeding: genotype-by-environment interaction, in: Goodman, R.M. (Ed.), Encyclopedia of Plant and Crop Science. Marcel-Dekker, New York, pp. 218–221.

Kao, C.H., Zeng, Z.B., Teasdale, R.D., 1999. Multiple interval mapping for quantitative trait loci. Genetics 152, 1203–1216. doi:10.1534/genetics.108.099028

Karki, S., Rizal, G., Quick, W.P., 2013. Improvement of photosynthesis in rice (Oryza sativa L.) by inserting the C4 pathway. Rice 6, 28. doi:10.1186/1939-8433-6-28

Karlsson, E.K., Baranowska, I., Wade, C.M., Salmon Hillbertz, N.H.C., Zody, M.C., Anderson, N., Biagi, T.M., Patterson, N., Pielberg, G.R., Kulbokas, E.J., Comstock, K.E., Keller, E.T., Mesirov, J.P., von Euler, H., Kämpe, O., Hedhammar, A., Lander, E.S., Andersson, G., Andersson, L., Lindblad-Toh, K., 2007. Efficient mapping of mendelian traits in dogs through genome-wide association. Nat. Genet. 39, 1321–1328. doi:10.1038/ng.2007.10

Kaya, Y., Akcura, M., Taner, S., 2006. GGE-biplot analysis of multi-environment yield trials in bread wheat. Turk. J. Agric. For. 30, 325–337.

Kearsey, M.J., Farquhar, A.G.L., 1998. QTL analysis in plants; where are we now? Heredity (Edinb). 80, 137–142. doi:10.1046/j.1365-2540.1998.00500.x

Kearsey, M.J., Pooni, H.S., 1996. The Genetical Analysis of Quantitative Traits. Chapman and Hall, London.

Kempton, R.A., Fox, P.N., 1997. Statistical methods for plant variety evaluation. Chapman and Hall, London.

Khan, M.H., Dar, Z.A., Dar, S.A., 2015. Breeding strategies for improving rice yield-A review. Agric. Sci. 6, 467–478. doi:10.4236/as.2015.65046

Khush, G.S., 1984. Terminology for Rice Growing Environments. International Rice Research Institute, Philippines.

Khush, G.S., 2005. What it will take to feed 5.0 billion rice consumers in 2030. Plant Mol. Biol. 59, 1–6. doi:10.1007/s11103-005-2159-5

Khush, G.S., Bacalangco, E., Ogawa, T., 1990. A new gene for resistance to bacterial blight from O. longistaminata. Rice Genet. Newsl. 7, 121–122.

Khush, G.S., Jena, K.K., 2009. Current status and future prospects for research on blast resistance in rice (Oryza sativa L.), in: Wang, G.L., Valent, B. (Eds.), Advances in Genetics, Genomics and Control of Rice Blast Disease. Springer, Dordrecht, p. pp 1–10.

Kim, H., Lee, K., Hwang, H., Bhatnagar, N., Kim, D., Yoon, I.S., Byun, M., Kim, S.T., Jung,

K., Kim, B., 2014. Overexpression of PYL5 in rice enhances drought tolerance , inhibits growth , and modulates gene expression. J. Exp. Bot. 65, 453–464. doi:10.1093/jxb/ert397

Kim, S.-R., Ramos, J., Ashikari, M. Virk, P.S., Torres, E.A., Nissila, E., Hechanova, S.L., Mauleon, R., Jena, K. K. 2016. Development and validation of allele-specific SNP/indel markers for eight yield-enhancing genes using whole-genome sequencing strategy to increase yield potential of rice, *Oryza sativa L*. Rice 9,12

Kloth, K.J., Thoen, M.P.M., Bouwmeester, H.J., Jongsma, M.A., Dicke, M., 2012. Association mapping of plant resistance to insects. Trends Plant Sci. 17, 311–319. doi:10.1016/j.tplants.2012.01.002

Kobayashi, A., Tomita, K., 2008. QTL detection for stickiness of cooked rice using recombinant inbred lines derived from crosses between japonica rice cultivars. Breed. Sci. 58, 419–426. doi:10.1270/jsbbs.58.419

Kobayashi, K., Maekawa, M., Miyao, A., Hirochika, H., Kyozuka, J., 2010. PANICLE PHYTOMER2 (PAP2), encoding a SEPALLATA subfamily MADS-box protein, positively controls spikelet meristem identity in rice. Plant cell Physiol. 51, 47–57. doi:10.1093/pcp/pcp166

Kojima, S., Takahashi, Y., Kobayashi, Y., Monna, L., Sasaki, T., Araki, T., Yano, M., 2002. Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. Plant Cell Physiol. 43, 1096–1105.

Kole, C., Muthamilarasan, M., Henry, R., Edwards, D., Sharma, R., Abberton, M., Batley, J., Bentley, A., Blakeney, M., Bryant, J., Cai, H., Cakir, M., Cseke, L.J., Cockram, J., de Oliveira, A.C., Pace, C. De, Dempewolf, H., Ellison, S., Gepts, P., Greenland, A., Hall, A., Hori, K., Hughes, S., Humphreys, M.W., Iorizzo, M., Ismail, A.M., Marshall, A., Mayes, S., Nguyen, H.T., Ogbonnaya, F.C., Ortiz, R., Paterson, A.H., Simon, P., Tohme, J., Tuberosa, R., Valliyodan, B., Varshney, R.K., Wullschleger, S.D., Yano, M., Manoj Prasad, 2015. Application of genomics-assisted breeding for generation of climate resilient crops: progress and prospects. Front. Plant Sci. 6, 563. doi:10.3389/fpls.2015.00563

Kumar, A., Dixit, S., Ram, T., Yadaw, R.B., Mishra, K.K., Mandal, N.P., 2014. Breeding high-yielding drought-tolerant rice: genetic variations and conventional and molecular approaches. J. Exp. Bot. 65, 6265–6278. doi:10.1093/jxb/eru363

Kumar, S., AlAbed, D., Worden, A., Novak, S., Wu, H., Ausmus, C., Beck, M., Robinson, H., Minnicks, T., Hemingway, D., Lee, R., Skaggs, N., Wang, L., Marri, P., Manju Gupta,

2015. A modular gene targeting system for sequential transgene stacking in plants. J. Biotechnol. 207, 12–20. doi:10.1016/j.jbiotec.2015.04.006

Kumar, V., Singh, A., Mithra, S.V.A., Krishnamurthy, S.L., Parida, S.K., Jain, S., Tiwari, K.K., Kumar, P., Rao, A.R., Sharma, S.K., Khurana, J.P., Singh, N.K., Mohapatra, T., 2015. Genome-wide association mapping of salinity tolerance in rice (Oryza sativa). DNA Res. 22, 133–145. doi:10.1093/dnares/dsu046

Kunert, A., Naz, A.A., Dedeck, O., Pillen, K., Léon, J., 2007. AB-QTL analysis in winter wheat: I. Synthetic hexaploid wheat (T. turgidum ssp. dicoccoides × T. tauschii) as a source of favourable alleles for milling and baking quality traits. Theor. Appl. Genet. 115, 683–695. doi:10.1007/s00122-007-0600-7

Lafitte, R., Blum, A., Atlin, G., 2003. Breeding Rice for Drought-Prone Environments. International Rice Research Institute, Los Baños (Philippines).

Lander, E.S., Botstein, D., 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121, 185–199.

Lander, E.S., Schork, N.J., 1994. Genetic dissection of complex traits. Nat. Genet. 265, 2037–2048.

Lao, O., van Duijn, K., Kersbergen, P., de Knijff, P., Kayser, M., 2006. Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. Am. J. Hum. Genet. 78, 680–690.

Leegood, R.C., 2013. Strategies for engineering C4 photosynthesis. J. Plant Physiol. 170, 378–388. doi:10.1016/j.jplph.2012.10.011

Lemaire, G., Gastal, F., 1997. N uptake and distribution in plant canopies, in: Lemaire, G. (Ed.), Diagnosis of the Nitrogen Status in Crops. Springer-Verlag, Berlin Heidelberg, pp. 3–44.

Li, H., Ribaut, J.M., Li, Z., Wang, J., 2008. Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. Theor. Appl. Genet. 116, 243–260. doi:10.1007/s00122-007-0663-5

Li, J., Thomson, M., McCouch, S.R., 2004. Fine mapping of a grain-weight quantitative trait locus in the pericentromeric region of rice chromosome 3. Genetics 168, 2187–2195. doi:10.1534/genetics.104.034165

Li, J., Wang, J., Zeigler, R.S., 2014. The 3 , 000 rice genomes project: new opportunities and challenges for future rice research. Gigascience 3, 8. doi:10.1186/2047-217X-3-8

Li, S., Qian, Q., Fu, Z., Zeng, D., Meng, X., Kyozuka, J., Maekawa, M., Zhu, X., Zhang, J., Li, J., Wang, Y., 2009. Short panicle1 encodes a putative PTR family transporter and

determines rice panicle size. Plant J. 58, 592–605. doi:10.1111/j.1365-313X.2009.03799.x

Li, T., Liu, B., Spalding, M.H., Weeks, D.P., Yang, B., 2012. High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat. Biotechnol. 30, 390–392. doi:10.1038/nbt.2199

Li, X., Qian, Q., Fu, Z., Wang, Y., Xiong, G., Zeng, D., Wang, X., Liu, X., Teng, S., Hiroshi, F., Yuan, M., Luo, D., Han, B., Li, J., 2003. Control of tillering in rice. Nature 422, 618–621. doi:10.1038/nature01518

Li, X., Quigg, R.J., Zhou, J., Xu, S., Masinde, G., Mohan, S., Baylink, D.J., 2006. A critical evaluation of the effect of population size and phenotypic measurement on QTL detection and localization using a large F2 murine mapping population. Genet. Mol. Biol. 29, 166–173. doi:10.1007/11760146_15

Li, X., Yan, W., Agrama, H., Jia, L., Jackson, A., Moldenhauer, K., Yeater, K., McClung, A., Wu, D., 2012. Unraveling the complex trait of harvest index with association mapping in rice (Oryza sativa L.). PLoS ONE 7(1), e29350. doi:10.1371/journal.pone.0029350

Li, X., Yan, W., Agrama, H., Jia, L., Shen, X., Jackson, A., Moldenhauer, K., Yeater, K., McClung, A., Wu, D., 2011. Mapping QTLs for improving grain yield using the USDA rice mini-core collection. Planta 234, 347–361. doi:10.1007/s00425-011-1405-0

Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., Shao, D., Xu, C., Li, X., Xiao, J., He, Y., Zhang, Q., 2011. Natural variation in GS5 plays an important role in regulating grain size and yield in rice. Nat. Genet. 43, 1266–1269. doi:10.1038/ng.977

Li, Y., Hallerman, E.M., Liu, Q., Wu, K., Peng, Y., 2015. The development and status of Bt rice in China. Plant Biotechnol J. Online. doi:10.1111/pbi.12464

Li, Z., Fu, B., Gao, Y., Xu, J., Ali, J., Lafitte, H.R., Jiang, Y., Rey, J.D., Vijayakumar, C.H.M., Maghirang, R., Zheng, T., Zhu, L., 2005. Genome-wide introgression lines and their use in genetic and molecular dissection of complex phenotypes in rice (Oryza sativa L.). Plant Mol. Biol. 59, 33–52. doi:10.1007/s11103-005-8519-3

Li, Z., Mu, P., Li, C., Zhang, H., Li, Z., Gao, Y., Wang, X., 2005. QTL mapping of root traits in a doubled haploid population from a cross between upland and lowland japonica rice in three environments. Theor. Appl. Genet. 110, 1244–1252. doi:10.1007/s00122-005-1958-z

Li, Z., Zhu, Y., 1988. Rice male sterile cytoplasm and fertility restorationin., in: Hybrid Rice. InternationalRice Research Institute, Manila, Philippines, pp. 85–102.

Lin, H., Ashikari, M., Yamanouchi, U., Sasaki, T., Yano, M., 2002. Identification and

characterization of a quantitative trait locus, Hd9, controlling heading date in rice. Breed. Sci. 52, 35–41. doi:10.1270/jsbbs.52.35

Lin, H., Wang, R., Qian, Q., Yan, M., Meng, X., Fu, Z., Yan, C., Jiang, B., Su, Z., Li, J., Wang, Y., 2009. DWARF27, an iron-containing protein required for the biosynthesis of strigolactones, regulates rice tiller bud outgrowth. Plant Cell 21, 1512–1525. doi:10.1105/tpc.109.065987

Linh, L.-H., Hang, N.-T., Jin, F.-X., Kang, K.-H., Lee, Y.-T., Kwon, S.-J., Ahn, S.-N., 2008. Introgression of a quantitative trait locus for spikelets per panicle from Oryza minuta to the O . sativa cultivar Hwaseongbyeo. Plant Breed. 267, 262–267. doi:10.1111/j.1439-0523.2007.01462.x

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S., Zhang, Z., 2012. GAPIT: Genome association and prediction integrated tool. Bioinformatics 28, 2397–2399. doi:10.1093/bioinformatics/bts444

Lipshutz, R.J., Fodor, S.P.A., Gingeras, T.R., Lockhart, D.J., 1999. High density synthetic oligonucleotide arrays. Nat. Genet. 21, 20–24.

Liu, K., Muse, V.S., 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21, 2128–2129. doi:10.1093/bioinformatics/bti282

Liu, N., Zhao, H., 2006. A non-parametric approach to population structure inference using multilocus genotypes. Hum. Genomics 2, 353–364. doi:10.1186/1479-7364-2-6-353

Liu, T., Mao, D., Zhang, S., Xu, C., Xing, Y., 2009. Fine mapping SPP1, a QTL controlling the number of spikelets per panicle, to a BAC clone in rice (Oryza sativa). Theor. Appl. Genet. 118, 1509–1517. doi:10.1007/s00122-009-0999-0

Liu, T., Shao, D., Kovi, M.R., Xing, Y., 2010. Mapping and validation of quantitative trait loci for spikelets per panicle and 1,000-grain weight in rice (Oryza sativa L.). Theor. Appl. Genet. 120, 933–942. doi:10.1007/s00122-009-1222-z

Liu, T., Yu, T., Xing, Y., 2013. Identification and validation of a yield-enhancing QTL cluster in rice (Oryza sativa L.) Euphytica 192, 145-153. doi: 10.1007/s10681-013-0929-8

Liu, W., Wu, C., Fu, Y., Hu, G., Si, H., Zhu, L., Luan, W., He, Z., Sun, Z., 2009. Identification and characterization of HTD2: A novel gene negatively regulating tiller bud outgrowth in rice. Planta 230, 649–658. doi:10.1007/s00425-009-0975-6

Lu, L., Yan, W., Xue, W., Shao, D., Xing, Y., 2012. Evolution and association analysis of Ghd7 in rice. PLoS ONE 7(5), e34021. doi:10.1371/journal.pone.0034021

Luo, X., Ji, S., Yuan, P., Lee, H., Kim, D., Balkunde, S., Kang, J.-W., Ahn, S.-N., 2013. QTL

mapping reveals a tight linkage between QTLs for grain weight and panicle spikelet number in rice. Rice 6, 33.

Ma, L., Yang, C., Zeng, D., Cai, J., Li, X., Ji, Z., Xia, Y., Qian, Q., Bao, J., 2009. Mapping QTLs for heading synchrony in a doubled haploid population of rice in two environments. J. Genet. Genomics 36, 297–304. doi:10.1016/S1673-8527(08)60118-6

Mackay, I., Powell, W., 2007. Methods for linkage disequilibrium mapping in crops. Trends Plant Sci. 12, 57–63. doi:10.1016/j.tplants.2006.12.001

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., Church, G.M., 2013. RNA-guided human genome engineering via Cas9. Science 339, 823–825. doi:10.1126/science

Mallik, S., Mandal, B.K., Sen, S.N., Sarkarung, S. 2002. Shuttle-breeding: An effective tool for rice varietal improvement in rainfed lowland ecosystems in eastern India. Current Sci. 83,1097-1102.

Marchini, J., Howie, B., 2010. Genotype imputation for genome-wide association studies. Nat. Rev. Genet. 11, 499–511. doi:10.1038/nrg2796

Martinez, V., Thorgaard, G., Robison, B., Sillanpää, M.J., 2005. An application of Bayesian QTL mapping to early development in double haploid lines of rainbow trout including environmental effects. Genet. Res. 86, 209–221. doi:10.1017/S0016672305007871

Mather, K.A., Caicedo, A.L., Polato, N.R., Olsen, K.M., McCouch, S., Purugganan, M.D., 2007. The extent of linkage disequilibrium in rice (Oryza sativa L.). Genetics 177, 2223–2232. doi:10.1534/genetics.107.079616

McCouch, S.R., Sweeney, M., Li, J., Jiang, H., Thomson, M., Septiningsih, E., Edwards, J., Moncada, P., Xiao, J., Garris, A., Tai, T., Martinez, C., Tohme, J., Sugiono, M., McClung, A., Yuan, L.P., Ahn, S.-N., 2007. Through the genetic bottleneck: O. rufipogon as a source of trait-enhancing alleles for O. sativa. Euphytica 154, 317–339. doi:10.1007/s10681-006-9210-8

McNally, K.L., Childs, K.L., Bohnert, R., Davidson, R.M., Zhao, K., Ulat, V.J., Zeller, G., Clark, R.M., Hoen, D.R., Bureau, T.E., Stokowski, R., Ballinger, D.G., Frazer, K. a, Cox, D.R., Padhukasahasram, B., Bustamante, C.D., Weigel, D., Mackill, D.J., Bruskiewich, R.M., Rätsch, G., Buell, C.R., Leung, H., Leach, J.E., 2009. Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc. Natl. Acad. Sci. U. S. A. 106, 12273–12278. doi:10.1073/pnas.0900992106

Meuwissen, T.H., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–1829.

Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., Wan, J., Gu, H., Qu, L.J., 2013. Targeted mutagenesis in rice using CRISPR-Cas system. Cell Res. 23, 1233–1236. doi:10.1038/cr.2013.123

Mikami, I., Uwatoko, N., Ikeda, Y., Yamaguchi, J., Hirano, H.Y., Suzuki, Y., Y. Sano, 2008. Allelic diversification at the wx locus in landraces of Asian rice. Theor. Appl. Genet. 116, 979–989. doi:10.1007/s00122-008-0729-z

Mishra, B., 2002. Genetic diversity in rice production, Case studies from Brazil, India and Nigeria., in: Nguyen, V.N. (Ed.), Genetic Diversity in Rice Production, Case Studies from Brazil, India and Nigeria. Food and Agriculture Organization of the United Nations (FAO), Rome, Italy, pp. 37–91.

Mitrovic, B., Stanisavljevi, D., Treski, S., Stojakovic, M., Ivanovic, M., Bekavac, G., Rajkovic, M., 2012. Evaluation of experimental maize hybrids tested in multi-location trials using AMMI and GGE biplot analysis. Turkish J. F. Crop. 17, 35–40.

Miura, K., Ashikari, M., Matsuoka, M., 2011. The role of QTLs in the breeding of high-yielding rice. Trends Plant Sci. 16, 319–326. doi:10.1016/j.tplants.2011.02.009

Miura, K., Ikeda, M., Matsubara, A., Song, X., Ito, M., Asano, K., Matsuoka, M., 2010. OsSPL14 promotes panicle branching and higher grain productivity in rice. Nat. Genet. 42, 545–549. doi:10.1038/ng.592

Mohammadi, M., Blake, T.K., Budde, A.D., Chao, S., Hayes, P.M., Horsley, R.D., Obert, D.E., Ullrich, S.E., Smith, K.P., 2015. A genome-wide association study of malting quality across eight U.S. barley breeding programs. Theor. Appl. Genet. 128, 705–721. doi:10.1007/s00122-015-2465-5

Molina, J., Sikora, M., Garud, N., Flowers, J.M., Rubinstein, S., Reynolds, A., Huang, P., Jackson, S., Schaal, B. a, Bustamante, C.D., Boyko, A.R., Purugganan, M.D., 2011. Molecular evidence for a single evolutionary origin of domesticated rice. Proc. Natl. Acad. Sci. U. S. A. 108, 8351–8356. doi:10.1073/pnas.1104686108

Moncada, P., Martínez, C.P., Borrero, J., Chatel, M., Jr, H.G., Guimaraes, E., Tohme, J., McCouch, S.R., 2001. Quantitative trait loci for yield and yield components in an Oryza sativa × Oryza rufipogon BC_2F_2 population evaluated in an upland environment. Theor. Appl. Genet. 102, 41–52. doi:10.1007/s001220051616

Monna, L., Kitazawa, N., Yoshino, R., Suzuki, J., Masuda, H., Maehara, Y., Tanji, M., Sato, M., Nasu, S., Minobe, Y., 2002. Positional cloning of rice semidwarfing gene, sd-1: Rice "green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis. DNA Res. 9, 11–17. doi:10.1093/dnares/9.1.11

Montalban, R., Destro, D., Silva, E.F., Montano, J.C., 1998. Genetic base of Brazilian upland rice cultivars. J. Genet. Breed 52, 203–209.

Mountain, J.L., Cavalli-sforza, L.L., 1997. Multilocus genotypes, a tree of individuals, and human evolutionary history. Am. J. Hum. Genet. 61, 705–718. doi:10.1086/515510

Muller, B.U., Kleinknecht, K., Mohring, J., Piepho, H.P., 2010. Comparison of spatial models for sugar beet and barley trials. Crop Sci. 50, 794–802. doi:10.2135/cropsci2009.03.0153

Murray, M.G., Thompson, W.F., 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 8, 4321–4326. doi:10.1093/nar/8.19.4321

Myles, S., Peiffer, J., Brown, P.J., Ersoz, E.S., Zhang, Z., Costich, D.E., Buckler, E.S., 2009. Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21, 2194–202. doi:10.1105/tpc.109.068437

Nachimuthu, V.V., Muthurajan, R., Duraialaguraja, S., Sivakami, R., Pandian, B.A., Ponniah, G., Gunasekaran, K., Swaminathan, M., K K, S., Sabariappan, R., 2015. Analysis of population structure and genetic diversity inrRice germplasm using SSR markers: An initiative towards association mapping of agronomic traits in Oryza sativa. Rice 8, 30. doi:10.1186/s12284-015-0062-5

Nakaya, A., Isobe, S.N., 2012. Will genomic selection be a practical method for plant breeding? Ann. Bot. 110, 1303–1316. doi:10.1093/aob/mcs109

Narasimhamoorthy, B., Gill, B.S., Fritz, A.K., Nelson, J.C., Brown-Guedira, G.L., 2006. Advanced backcross QTL analysis of a hard winter wheat × synthetic wheat population. Theor. Appl. Genet. 112, 787–796. doi:10.1007/s00122-005-0159-0

Nawaz, Z., Kakar, K.U., Li, X., Li, S., Zhang, B., Shou, H., Shu, Q., 2015. Genome-wide association mapping of quantitative trait loci (QTLs) for contents of eight elements in brown rice (Oryza sativa L.). J. Agric. Food Chem. 63, 8008–8016. doi:10.1021/acs.jafc.5b01191

Negrão, S., Almadanim, M.C., Pires, I.S., Abreu, .IA., Maroco, J., Courtois, B., Gregorio, G.B., McNally, K.L., Oliveira, M.M., 2013. New allelic variants found in key rice salt-tolerance genes: an association study. Plant Biotechnol. J. 11, 87–100. doi:10.1111/pbi.12010

Nei, M., Li, W.H., 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. U. S. A. 76, 5269–5273. doi:10.1073/pnas.76.10.5269

Newell-mcgloughlin, M., 2008. Nutritionally Improved Agricultural Crops. Plant Physiolgoy

147, 939–953. doi:10.1104/pp.108.121947

Nguyen, N., Ferrero, A., 2006. Meeting the challenges of global rice production. Paddy Water Environ. 4, 1–9. doi:10.1007/s10333-005-0031-5

Niño-Liu, D.O., Ronald, P.C., Bogdanove, A.J., 2006. Xanthomonas oryzae pathovars: Model pathogens of a model crop. Mol. Plant Pathol. 7, 303–324. doi:10.1111/j.1364-3703.2006.00344.x

Obara, M., Sato, T., Sasaki, S., Kashiba, K., Nagano, A., Nakamura, I., Ebitani, T., Yano, M., Yamaya, T., 2004. Identification and characterization of a QTL on chromosome 2 for cytosolic glutamine synthetase content and panicle number in rice. Theor. Appl. Genet. 110, 1–11. doi:10.1007/s00122-004-1828-0

Obara, M., Tamura, W., Ebitani, T., Yano, M., Sato, T., Yamaya, T., 2010. Fine-mapping of qRL6.1, a major QTL for root length of rice seedlings grown under a wide range of NH_4^+ concentrations in hydroponic conditions. Theor. Appl. Genet. 121, 535–547. doi:10.1007/s00122-010-1328-3

Ogura, T., Busch, W., 2015. From phenotypes to causal sequences: using genome wide association studies to dissect the sequence basis for variation of plant development. Plant Biol. 23, 98–108.

Oh, J.-M., Yoon, D.-B., Ahn, S.-N., 2010. Fine mapping of grain weight QTLs using near isogenic lines from a Cross between Oryza sativa and O. grandiglumis. J. Crop Sci. Biotechnol. 13, 7–12. doi:10.1007/s12892-010-0015-x

Ohsumi, A., Takai, T., Ida, M., Yamamoto, T., Arai-Sanoh, Y., Yano, M., Ando, T., Kondo, M., 2011. Evaluation of yield performance in rice near-isogenic lines with increased spikelet number. Field Crops Res. 120, 68–75. doi:10.1016/j.fcr.2010.08.013

Oliva, N., Chadha-Mohanty, P., Poletti, S., Abrigo, E., Atienza, G., Torrizo, L., Garcia, R., Jr., C.D., Poncio, M.A., Balindong, J., Manzanilla, M., Montecillo, F., Zaidem, M., Barry, G., Hervé, P., Shou, H., Slamet-Loedin, I.H., 2014. Large-scale production and evaluation of marker-free indica rice IR64 expressing phytoferritin genes. Mol. Breed. 33, 23–37. doi:10.1007/s11032-013-9931-z

Onogi, A., Watanabe, M., Mochizuki, T., Hayashi, T., Nakagawa, H., Hasegawa, T., Iwata, H., 2016. Toward integration of genomic selection with crop modelling: the development of an integrated approach to predicting rice heading dates. Theor. Appl. Genet. Online. doi:10.1007/s00122-016-2667-5

Ookawa, T., Hobo, T., Yano, M., Murata, K., Ando, T., Miura, H., Asano, K., Ochiai, Y., Ikeda, M., Nishitani, R., Ebitani, T., Ozaki, H., Angeles, E.R., Hirasawa, T., Matsuoka,

M., 2010. New approach for rice improvement using a pleiotropic QTL gene for lodging resistance and yield. Nat. Commun. 1, 132. doi:10.1038/ncomms1132

Ordonez Jr., S.A., Silva, J., Oard, J.H., 2010. Association mapping of grain quality and flowering time in elite japonica rice germplasm. J. Cereal Sci. 51, 337–343. doi:10.1016/j.jcs.2010.02.001

Ouk, M., Basnayake, J., Tsubo, M., Fukai, S., Fischer, K.S., Kang, S., Men, S., Thun, V., Cooper, M., 2007. Genotype-by-environment interactions for grain yield associated with water availability at flowering in rainfed lowland rice. Field Crops Res. 101, 145–154. doi:10.1016/j.fcr.2006.10.003

Paine, J.A., Shipton, C.A., Chaggar, S., Howells, R.M., Kennedy, M.J., Vernon, G., Wright, S.Y., Hinchliffe, E., Adams, J.L., Silverstone, A.L., Drake, R., 2005. Improving the nutritional value of Golden Rice through increased pro-vitamin A content. Nat. Biotechnol. 23, 482–488. doi:10.1038/nbt1082

Papademetriou, M.K., Dent, F.J., Herath, E.M., 2000. Bridging the rice yield gap in the Asian-Pacific region. Food and Agriculture Organization of the United Nations regional office for Asia and the Pacific, Bankok, Thailand.

Pasam, R.K., Sharma, R., Malosetti, M., van Eeuwijk, F.A., Haseneyer, G., Kilian, B., Graner, A., 2012. Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC Plant Biol. 12, 16. doi:10.1186/1471-2229-12-16

Paterson, A.H., Freeling, M., Sasaki, T., 2005. Grains of knowledge: Genomics of model cereals. Genome Res. 15, 1643–1650. doi:10.1101/gr.3725905

Pauli, D., Muehlbauer, G., Smith, K., Cooper, B., 2014. Association mapping of agronomic QTLs in US spring barley breeding germplasm. Plant Genome 7, 1–15. doi:10.3835/plantgenome2013.11.0037

Peng, S., Cassman, K.G., Virmani, S.S., Sheehy, J., Khush, G.S., 1999. Yield potential trends of tropical rice since the release of IR8 and the challenge of increasing rice yield potential. Crop Sci. 39, 1552–1559. doi:10.2135/cropsci1999.3961552x

Peng, S., Huang, J., Cassman, K.G., Laza, R.C., Visperas, R.M., Khush, G.S., 2010. The importance of maintenance breeding: A case study of the first miracle rice variety-IR8. Field Crops Res. 119, 342–347. doi:10.1016/j.fcr.2010.08.003

Peng, S., Khush, G.S., 2003. Four Decades of Breeding for Varietal Improvement Irrigated Lowland Rice in the International Rice. Plant Prod. Sci. 6, 157–164.

Peng, S., Khush, G.S., Cassman, K.G., 1994. Evaluation of a new plant ideotype for increased yield potential., in: Cassman, K.G. (Ed.), Breaking the Yield Barrier:

Proceedings of a Workshop on Rice Yield Potential in Favourable Environments. International Rice Research Institute, pp. 5–20.

Peng, S., Laza, R.C., Visperas, R.M., Sanico, A.L., Cassman, K.G., Khush, G.S., 2000. Grain yield of rice cultivars and lines developed in the Philippines since 1966. Crop Sci. 40, 307–314.

Peters, J.L., Cnudde, F., Gerats, T., 2003. Forward genetics and map-based cloning approaches. Trends Plant Sci. 8, 484–491. doi:10.1016/j.tplants.2003.09.002

Petolino, J.F., Kumar, S., 2016. Transgenic trait deployment using designed nucleases. Plant Biol. J. 14, 503–509. doi:10.1111/pbi.12457

Phung, N.T.P., Mai, C.D., Mournet, P., Frouin, J., Droc, G., Ta, N.K., Jouannic, S., Lê, L.T., Do, V.N., Gantet, P., Courtois, B., 2014. Characterization of a panel of Vietnamese rice varieties using DArT and SNP markers for association mapping purposes. BMC Plant Biol. 14, 1–16. doi:10.1186/s12870-014-0371-7

Piao, R., Jiang, W., Ham, T.-H., Choi, M.-S., Qiao, Y., Chu, S.-H., Park, J.-H., Woo, M.-O., Jin, Z., An, G., Lee, J., Koh, H.-J., 2009. Map-based cloning of the ERECT PANICLE 3 gene in rice. Theor. Appl. Genet. 119, 1497–1506. doi:10.1007/s00122-009-1151-x

Piepho, H., Mohring, J., Melchinger, A., Buchse, A., 2008. BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161, 209–228.

Piepho, H.P., 2001. A quick method for computing approximate thresholds for quantitative trait loci detection. Genetics 157, 425–432.

Platt, A., Horton, M., Huang, Y.S., Li, Y., Anastasio, A.E., Mulyati, N.W., Agren, J., Bossdorf, O., Byers, D., Donohue, K., Dunning, M., Holub, E.B., Hudson, A., Le Corre, V., Loudet, O., Roux, F., Warthmann, N., Weigel, D., Rivero, L., Scholl, R., Nordborg, M., Bergelson, J., Borevitz, J.O., 2010. The scale of population structure in Arabidopsis thaliana. PLoS Genet. 6(2), e1000843. doi:10.1371/journal.pgen.1000843

Platten, J.D., Egdane, J.A., Ismail, A.M., 2013. Salinity tolerance, Na^+ exclusion and allele mining of HKT1;5 in Oryza sativa and O. glaberrima: many sources, many genes, one mechanism? BMC Plant Biol. 13, 32. doi:10.1186/1471-2229-13-32

Pozniak, C.J., Clarke, J.M., Clarke, F.R., 2012. Potential for detection of marker-trait associations in durum wheat using unbalanced, historical phenotypic datasets. Mol. Breed. 30, 1537–1550. doi:10.1007/s11032-012-9737-4

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909. doi:10.1038/ng1847

Pritchard, J.K., Rosenberg, N.A., 1999. Use of unlinked genetic markers to detect population stratification in asscociation studies. Am. J. Hum. Genet. 65, 220–228.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000a. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Pritchard, J.K., Stephens, M., Rosenberg, N.A., Donnelly, P., 2000b. Association mapping in structured populations. Am. J. Hum. Genet. 67, 170–181. doi:10.1086/302959

Purcell, S., Neale, B., Todd-brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., Bakker, P.I.W. De, Daly, M.J., Sham, P.C., 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. doi:10.1086/519795

Purcell, S., Sham, P., 2004. Properties of structured association approaches to detecting population stratification. Hum. Hered. 8, 93–107. doi:10.1159/000083030

Pusadee, T., Jamjod, S., Chang, Y.-C., Rerkasem, B., Schaal, B.A., 2009. Genetic structure and isolation by distance in a landrace of Thai rice. Proc. Natl. Acad. Sci. U.S.A. 106, 13880–13885. doi:10.1073/pnas.0906720106

Qian, Q., Guo, L.B., Yang, C.D., 2007. Gene-Based Designed Breeding in Rice (in Chinese). Science Press, Beijing.

Qiao, Y., Piao, R., Shi, J., Lee, S.-I., Jiang, W., Kim, B.-K., Lee, J., Han, L., Ma, W., Koh, H.-J., 2011. Fine mapping and candidate gene analysis of dense and erect panicle 3, DEP3, which confers high grain yield in rice (Oryza sativa L.). Theor. Appl. Genet. 122, 1439–1449. doi:10.1007/s00122-011-1543-6

R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rai, M., 2002. Genetic diversity in rice production, past contribution and the potential of utilization for sustainable rice production, in: Sustainable Rice Production for Food Security. Proceedings of the 20th Session of the International Rice Commission. Bangkok, Thailand, pp. 89–115.

Rakshit, S., Kanzaki, H., Matsumura, H., Rakshit, A., Fujibe, T., Okuyama, Y., Yoshida, K., Oli, M., Shenton, M., Utsushi, H., Mitsuoka, C., Abe, A., Kiuchi, Y., Terauchi, R., 2010. Use of TILLING for reverse and forward genetics of rice, in: Meksem, K., Kahl, G. (Eds.), The Handbook of Plant Mutation Screening: Mining of Natural and Induced Alleles. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

Ranawake, A.L., Amarasinghe, U.G.S., 2014. Relationship of yield and yield related traits of some traditional rice cultivars in Sri Lanka as described by correlation analysis. J. Sci.

Res. Reports 3, 2395–2403.

Rangel, P.N., Brondani, R.P.V., Rangel, P.H.N., Brondani, C., 2008. Agronomic and molecular characterization of introgression lines from the interspecific cross Oryza sativa (BG90-2) × Oryza glumaepatula. Genet. Mol. Res. 7, 184–195.

Redoña, E.D., Mackill, D.J., 1998. Quantitative trait locus analysis for rice panicle and grain characteristics. Theor. Appl. Genet. 96, 957–963. doi:10.1007/s001220050826

Remington, D.L., Ungerer, M.C., Purugganan, M.D., 2001. Map-based cloning of quantitative trait loci: progress and prospects, Genet. Res., Camb.

Romay, M.C., Millard, M.J., Glaubitz, J.C., Peiffer, J.A., Swarts, K.L., Casstevens, T.M., Elshire, R.J., Acharya, C.B., Mitchell, S.E., Flint-Garcia, S.A., McMullen, M.D., Holland, J.B., Buckler, E.S., Gardner, C.A., 2013. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. 14, R55. doi:10.1186/gb-2013-14-6-r55

Russell, J.R., Ellis, R.P., Thomas, W.T.B., Waugh, R., Provan, J., Booth, A., Fuller, J., Lawrence, P., Young, G., Powell, W., 2000. A retrospective analysis of spring barley germplasm development from "foundation genotypes" to currently successful cultivars. Mol. Breed. 6, 553–568. doi:10.1023/A:1011372312962

Saitou, N., Nei, M., 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425. doi:citeulike-article-id:93683

Samonte, S.O.P.B., Hernandez, J.E., 1990. Genotype × environment interactions of irrigated lowland rice I. Variance components and optimum allocation of resources. Philipp. J. Crop Sci. 15, 55–66.

Samonte, S.O.P.B., Hernandez, J.E., 1991. Genotype × environment interactions of irrigated lowland rice: II. Stability and adaptability analysis. Philipp. J. Crop Sci. 26, 7–14.

Sander, J.D., Joung, J.K., 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol. 32, 347–355. doi:10.1038/nbt.2842

Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., Kilian, A., 2011. Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. BMC Proc. 5(Suppl 7), P54. doi:10.1186/1753-6561-5-S7-P54

Santos, F. R., Pena, S. D., Epplen, J.T., 1993. Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. Hum. Genet. 90, 655–656.

Sarkar, R.K., Bhattacharjee, B., 2011. Rice genotypes with SUB1 QTL differ in submergence

tolerance, elongation ability during submergence, and re-generation growth at re-emergence. Rice 5, 7. doi:10.1007/s12284-011-9065-z

Sarkar, R.K., Panda, D., Reddy, J.N., Patnaik, S.S.C., Mackill, D.J., Ismail, A.M., 2009. Performance of submergence tolerant rice (Oryza sativa) genotypes carrying the Sub1 quantitative trait locus under stressed and non-stressed natural field conditions. Indian J. Agric. Sci. 79, 876–883.

Sasaki, A., Ashikari, M., Ueguchi-Tanaka, M., Itoh, H., Nishimura, A., SWapan, D., Ishiyams, K., Saito, T., Kobayashi, M., Kush, G.S., Kitano, H., Matsuoka, M., 2002. Green Revolution: A mutant gibberellin-synthesis gene in rice. Nature 416, 701–702. doi:10.1038/416701a

Satagopan, J.M., Yandell, B.S., Newton, M. a, Osborn, T.C., 1996. A bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. Genetics 144, 805–816. doi:10.1186/1753-6561-5-S3-S4

Seck, P.A., Diagne, A., Mohanty, S., Wopereis, M.C.S., 2012. Crops that feed the world 7: Rice. Food Secur. 4, 7–24. doi:10.1007/s12571-012-0168-1

Semagn, K., Bjørnstad, Å., Ndjiondjop, M.N., 2006. Principles, requirements and prospects of genetic mapping in plants. African J. Biotechnol. 5, 2569–2587.

Septiningsih, E.M., Pamplona, A.M., Sanchez, D.L., Neeraja, C.N., Vergara, G. V., Heuer, S., Ismail, A.M., Mackill, D.J., 2008. Development of submergence-tolerant rice cultivars: the Sub1 locus and beyond. Ann. Bot. 103, 151–160. doi:10.1093/aob/mcn206

Shan, J.-X., Zhu, M.-Z., Shi, M., Gao, J.-P., Lin, H.-X., 2009. Fine mapping and candidate gene analysis of spd6, responsible for small panicle and dwarfness in wild rice (Oryza rufipogon Griff.). Theor. Appl. Genet. 119, 827–836. doi:10.1007/s00122-009-1092-4

Shan, Q., Wang, Y., Chen, K., Liang, Z., LI, J., Zhang, Y., Zhang, K., Liu, J., VOytas, D.F., Zheng, X., Zhang, Y., Caixia Gao, 2013a. Rapid and Efficient Gene Modification in Rice and Brachypodium Using TALENs. Mol. Plant 6, 1365–1368. doi:10.1093/mp/sss162

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.-L., Gao, C., 2013b. Targeted genome modification of crop plants using a CRISPR-Cas system. Nat. Biotechnol. 31, 686–688. doi:10.1038/nbt.2652

Shan, Q., Zhang, Y., Chen, K., Zhang, K., Gao, C., 2015. Creation of fragrant rice by targeted knockout of the OsBADH2 gene using TALEN technology. Plant Biotechnol J. 13, 791–800. doi:10.1111/pbi.12312

Shao, G., Tang, S., Luo, J., Jiao, G., Wei, X., Tang, A., Wu, J., Zhuang, J., Hu, P., 2010.

Mapping of qGL7-2, a grain length QTL on chromosome 7 of rice. J. Genet. genomics 37, 523–531. doi:10.1016/S1673-8527(09)60072-2

Shao, Y., Jin, L., Zhang, G., Lu, Y., Shen, Y., Bao, J., 2011. Association mapping of grain color, phenolic content, flavonoid content and antioxidant capacity in dehulled rice. Theor. Appl. Genet. 122, 1005–1016. doi:10.1007/s00122-010-1505-4

Shi, W., Yang, Y., Chen, S., Xu, M., 2008. Discovery of a new fragrance allele and the development of functional markers for the breeding of fragrant rice varieties. Mol. Breed. 22, 185–192. doi:10.1007/s11032-008-9165-7

Shi, X., Wang, J., Bao, Y., Li, P., Xie, L., Huang, J., H., Z., 2010. Identification of the quantitative trait loci in japonica rice landrace Heikezijing responsible for broad-spectrum resistance to rice blast. Phytopathology 100, 822–829.

Shimamoto, K., Kyozuka, J., 2002. Rice as a model for comparative genomics of plants. Annu. Rev. Plant. Biol. 53, 399–419. doi:10.1146/annurev.arplant.53.092401.134447

Shimatani, Z., Nishizawa-yokoi, A., Endo, M., Toki, S., Terada, R., 2015. Positive-negative-selection-mediated gene targeting in rice. Front. Plant Sci. 5, 748. doi:10.3389/fpls.2014.00748

Shomura, A., Izawa, T., Ebana, K., Ebitani, T., Kanegae, H., Konishi, S., Yano, M., 2008. Deletion in a gene associated with grain size increased yields during rice domestication. Nat. Genet. 40, 1023–1028. doi:10.1038/ng.169

Shriver, M., Kennedy, G.C., Parra, E.J., Lawson, H.A., Sonpar, V., Huang, J., Akey, O., Jones, K.W., 2004. The genomic distribution of population substructurein four populations using 8,525 autosomal SNPs. Hum. Genomics 1, 274–286. doi:10.1186/1479-7364-1-4-274

Shu, Q., Ye, G., Cui, H., Cheng, X., Xiang, Y., Wu, D., Gao, M., Xia, Y., Hu, C., Sardana, R., Altosaar, I., 2000. Transgenic rice plants with a synthetic cry1Ab gene from Bacillus thuringiensis were highly resistant to eight lepidopteran rice pest species. Mol. Breed. 6, 433–439. doi:10.1023/A:1009658024114

Sillanpa, M.J., Arjas, E., 1999. Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. Genetics 151, 1605–1619.

Sillanpää, M.J., 2011. Overview of techniques to account for confounding due to population stratification and cryptic relatedness in genomic data association analyses. Heredity (Edinb). 106, 511–519. doi:10.1038/hdy.2010.91

Silva, G., Poirot, L., Galetto, R., Smith, J., Montoya, G., Duchateau, P., Pâques, F., 2011. Meganucleases and other tools for targeted genome engineering: perspectives and

challenges for gene therapy. Curr. Gene Ther. 11, 11–27. doi:10.2174/156652311794520111

Simmonds, N.W. (Ed.), 1979. Evolution of Crop Plants. Longman, London.

Singh, A.K., Kumar, R., Tripathi, A.K., Gupta, B.K., Pareek, A., Singla-pareek, S.L., 2015. Genome-wide investigation and expression analysis of Sodium/Calcium exchanger gene family in rice and Arabidopsis. Rice 8, 21. doi:10.1186/s12284-015-0054-5

Singh, N., Choudhury, D.R., Singh, A.K., Kumar, S., Srinivasan, K., Tyagi, R.K., Singh, N.K., Singh, R., 2013. Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. PLoS ONE 8(12), e84136. doi:10.1371/journal.pone.0084136

Singh, S., Mackill, D.J., Ismail, A.M., 2009. Responses of SUB1 rice introgression lines to submergence in the field: Yield and grain quality. Field Crops Res. 113, 12–23. doi:10.1016/j.fcr.2009.04.003

Smith, A.B., Cullis, B.R., Thompson, R., 2005. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. J. Agric. Sci. 143, 449–462.

Soller, M., Brody, T., 1976. On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor. Appl. Genet. 47, 35–39.

Song, X.-J., Huang, W., Shi, M., Zhu, M.-Z., Lin, H.-X., 2007. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat. Genet. 39, 623–630. doi:10.1038/ng2014

Soto-Cerda, B.J., Cloutier, S., 2012. Association mapping in plant genomes, in: Caliskan, M. (Ed.), Genetic Diversity in Plants. InTech, Rijeka.

Spielmeyer, W., Ellis, M.H., Chandler, P.M., 2002. Semidwarf (sd-1), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. Proc. Natl. Acad. Sci. U. S. A. 99, 9043–9048. doi:10.1073/pnas.132266399

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., Atlin, G., Jannink, J.-L., McCouch, S.R., 2015. Genomic selection and association mapping in rice (Oryza sativa): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genet. 11(2), e1004982. doi:10.1371/journal.

Spindel, J., Wright, M., Chen, C., Cobb, J., Gage, J., Harrington, S., Lorieux, M., Ahmadi, N., McCouch, S., 2013. Bridging the genotyping gap: Using genotyping by sequencing

(GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theor. Appl. Genet. 126, 2699–2716. doi:10.1007/s00122-013-2166-x

Stich, B., Maurer, H.P., Melchinger, A.E., Frisch, M., Heckenberger, M., Voort, J.R., Peleman, J., Sørensen, A.P., Reif, J.C., 2006. Comparison of linkage disequilibrium in elite european maize inbred lines using AFLP and SSR markers. Mol. Breed. 17, 217–226. doi:10.1007/s11032-005-5296-2

Stich, B., Melchinger, A.E., 2009. Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and Arabidopsis. BMC Genomics 10, 94. doi:10.1186/1471-2164-10-94

Stich, B., Melchinger, A.E., Frisch, M., Maurer, H.P., Heckenberger, M., Reif, J.C., 2005. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. Theor. Appl. Genet. 111, 723–730. doi:10.1007/s00122-005-2057-x

Stich, B., Möhring, J., Piepho, H.-P., Heckenberger, M., Buckler, E.S., Melchinger, A.E., 2008a. Comparison of mixed-model approaches for association mapping. Genetics 178, 1745–1754. doi:10.1534/genetics.107.079707

Stich, B., Piepho, H.-P., Schulz, B., Melchinger, A.E., 2008b. Multi-trait association mapping in sugar beet (Beta vulgaris L.). Theor. Appl. Genet. 117, 947–954. doi:10.1007/s00122-008-0834-z

Storey, J.D., 2002. A direct approach to false discovery rates. J. R. Stat. Soc. B 64, 479–498.

Storey, J.D., Tibshirani, R., 2003. Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. 100, 9440–9445.

Sun, C., Su, J., Hu, C., Yan, X., Jin, Y., Chen, Z., Guan, Q., Wang, Y., Zhong, D., Jansson, C., Wang, F., Schnu, A., 2015. Expression of barley SUSIBA2 transcription factor yields high-starch low-methane rice. Nature 523, 602–606. doi:10.1038/nature14673

Sun, X.-Y., Wu, K., Zhao, Y., Kong, F.-M., Han, G.-Z., Jiang, H.-M., Huang, X.-J., Li, R.-J., Wang, H.-G., Li, S.-S., 2009. QTL analysis of kernel shape and weight using recombinant inbred lines in wheat. Euphytica 165, 615–624. doi:10.1007/s10681-008-9794-2

Swain, D.K., Herath, S., Pathirana, A., Mittra, B.N., 2005. Rainfed Lowland and Floodprone Rice: A Critical Review on Ecology and Management Technology for Improving the Productivity in Asia. Role of Water Sciences in Transboundary River Basin Management, Thailand.

Swamy, B.P.M., Kumar, A., 2012. Sustainable rice yield in water-short drought-prone

environments: conventional and molecular approaches, in: Lee, T.S. (Ed.), Irrigation Systems and Practices in Changing Environments. InTech, pp. 149–168.

Tabor, H.K., Risch, N.J., Myers, R.M., 2002. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat. Rev. Genet. 3, 391–397. doi:10.1038/nrg796

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., Uemura, A., Utsushi, H., Tamiru, M., Takuno, S., Innan, H., Cano, L.M., Kamoun, S., Terauchi, R., 2013a. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J. 74, 174–183. doi:10.1111/tpj.12105

Takagi, H., Uemura, A., Yaegashi, H., Tamiru, M., Abe, A., Mitsuoka, C., Utsushi, H., Natsume, S., Kanzaki, H., Matsumura, H., Saitoh, H., Cano, L.M., Kamoun, S., Terauchi, R., 2013b. Methods MutMap-Gap: whole-genome resequencing of mutant F2 progeny bulk combined with de novo assembly of gap regions identifies the rice blast resistance gene Pii. New Phytol. 200, 276–283.

Takano-Kai, N., Jiang, H., Kubo, T., Sweeney, M., Matsumoto, T., Kanamori, H., Padhukasahasram, B., Bustamante, C., Yoshimura, A., Doi, K., McCouch, S., 2009. Evolutionary history of GS3, a gene conferring grain length in rice. Genetics 182, 1323–1334. doi:10.1534/genetics.109.103002

Takeda, T., Suwa, Y., Suzuki, M., Kitano, H., Ueguchi-Tanaka, M., Ashikari, M., Matsuoka, M., Ueguchi, C., 2003. The OsTB1 gene negatively regulates lateral branching in rice. Plant J. 33, 513–20.

Tan, C.J., Sun, Y.J., Xu, H.S., Yu, S.B., 2011. Identification of quantitative trait locus and epistatic interaction for degenerated spikelets on the top of panicle in rice. Plant Breed. 130, 177–184. doi:10.1111/j.1439-0523.2010.01770.x

Tanksley, S.D., McCouch, S.R., 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277, 1063–1066. doi:10.1126/science.277.5329.1063

Tanksley, S.D., Nelson, J.C., 1996. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor. Appl. Genet. 92, 191–203. doi:10.1007/BF00223376

Tariku, S., Lakew, T., Bitew, M., Asfaw, M., 2013. Genotype by environment interaction and grain yield stability analysis of rice (Oryza sativa L.) genotypes evaluated in north western Ethiopia. Net J. Agric. Sci. 1, 10–16.

Terao, T., Nagata, K., Morino, K., Hirose, T., 2010. A gene controlling the number of

primary rachis branches also controls the vascular bundle formation and hence is responsible to increase the harvest index and grain yield in rice. Theor. Appl. Genet. 120, 875–893. doi:10.1007/s00122-009-1218-8

Tester, M., Langridge, P., 2010. Breeding technologies to increase crop production in a changing world. Science 327, 818–822. doi:10.1126/science.1183700

The 3000 Rice Genomes Project, 2014. The 3, 000 rice genomes project. Gigascience 3, 7. doi:10.1186/2047-217X-3-7

Thomson, M.J., de Ocampo, M., Egdane, J., Rahman, M.A., Sajise, A.G., Adorada, D.L., Tumimbang-Raiz, E., Blumwald, E., Seraj, Z.I., Singh, R.K., Gregorio, G.B., Ismail, A.M., 2010. Characterizing the Saltol quantitative trait locus for salinity tolerance in rice. Rice 3, 148–160. doi:DOI 10.1007/s12284-010-9053-8

Thomson, M.J., Septiningsih, E.M., Suwardjo, F., Santoso, T.J., Silitonga, T.S., McCouch, S.R., 2007. Genetic diversity analysis of traditional and improved Indonesian rice (Oryza sativa L.) germplasm using microsatellite markers. Theor. Appl. Genet. 117, 559–568. doi:10.1007/s00122-006-0457-1

Thomson, M.J., Tai, T.H., McClung, A.M., Lai, X.-H., Hinga, M.E., Lobos, K.B., Xu, Y., Martinez, C.P., McCouch, S.R., 2003. Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between Oryza rufipogon and the Oryza sativa cultivar Jefferson. Theor. Appl. Genet. 107, 479–493. doi:10.1007/s00122-003-1270-8

Tian, F., Zhu, Z., Zhang, B., Tan, L., Fu, Y., Wang, X., Sun, C.Q., 2006. Fine mapping of a quantitative trait locus for grain number per panicle from wild rice (Oryza rufipogon Griff.). Theor. Appl. Genet. 113, 619–629. doi:10.1007/s00122-006-0326-y

Till, B.J., Cooper, J., Tai, T.H., Colowit, P., Greene, E.A., Henikoff, S., Comai, L., 2007. Discovery of chemically induced mutations in rice by TILLING. BMC Plant Biol. 7, 19. doi:10.1186/1471-2229-7-19

Townsend, J.A., Wright, D.A., Winfrey, R.J., Fu, F., Maeder, M.L., Joung, J.K., Voytas, D.F., 2009. High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature 459, 442–445. doi:10.1038/nature07845

Tsai, H., Howell, T., Nitcher, R., Missirian, V., Watson, B., Ngo, K.J., Lieberman, M., Fass, J., Uauy, C., Tran, R.K., Khan, A.A., Filkov, V., Tai, T.H., Dubcovsky, J., Comai, L., 2011. Discovery of rare mutations in populations: TILLING by sequencing. Plant Physiolgoy 156, 1257–1268. doi:10.1104/pp.110.169748

Umehara, M., Hanada, A., Yoshida, S., Akiyama, K., Arite, T., Takeda-Kamiya, N., Magome,

H., Kamiya, Y., Shirasu, K., Yoneyama, K., Kyozuka, J., Yamaguchi, S., 2008. Inhibition of shoot branching by new terpenoid plant hormones. Nature 455, 195–200. doi:10.1038/nature07272

van Orsouw, N.J., Hogers, R.C.J., Janssen, A., Yalcin, F., Snoeijers, S., Verstege, E., Schneiders, H., van der Poel, H., van Oeveren, J., Verstegen, H., van Eijk, M.J.T., 2007. Complexity reduction of polymorphic sequences (CRoPS^{TM}): A novel approach for large-scale polymorphism discovery in complex genomes. PLoS ONE 2(11), e1172. doi:10.1371/journal.pone.0001172

Van Verk, M.C., Hickman, R., Pieterse, C.M.J., Van Wees, S.C.M., 2013. RNA-Seq: revelation of the messengers. Trends Plant Sci. 18, 175–179. doi:10.1016/j.tplants.2013.02.001

Wade, L.J., McLaren, C.G., Criseno, L., Amarante, S.T., Sarawgi, A.K., Kumar, R., Bhamri, M.C., Sing, O.N., Ahmed, H.U. Rajatasereekul, S. Pornuraisanit, P., Boonwite, C., Harnpichivitaya, D., Sarkarung, S., 1997. Genotype-by-environment interactions: RLRRC Experience, in: Fukai, S., Cooper, M., Salisbury, J. (Eds.), Breeding Strategies for Rainfed Lowland Rice in Drought-Prone Environments.ACIAR Proceedings 77. pp. 115–125.

Wade, L.J., McLaren, C.G., Quintana, L., Harnpichitvitaya, D., Rajatasereekul, S., Sarawgi, A.K., Kumar, A., Ahmed, H.U., Singh, A.K., Rodriguez, R., Siopongco, J., Sarkarung, S., 1999. Genotype by environment interactions across diverse rainfed lowland rice environments. Field Crops Res. 64, 35–50. doi:10.1016/S0378-4290(99)00049-0

Wan, X., Weng, J., Zhai, H., Wang, J., Lei, C., Liu, X., Guo, T., Jiang, L., Su, N., Wan, J., 2008. Quantitative trait loci (QTL) analysis for rice grain width and fine mapping of an identified QTL allele gw-5 in a recombination hotspot region on chromosome 5. Genetics 179, 2239–2252. doi:10.1534/genetics.108.089862

Wang, C., Chen, S., Yu, S., 2011. Functional markers developed from multiple loci in GS3 for fine marker-assisted selection of grain length in rice. Theor. Appl. Genet. 122, 905–913. doi:10.1007/s00122-010-1497-0

Wang, C., Yang, Y., Yuan, X., Xu, Q., Feng, Y., Yu, H., Wang, Y., Wei, X., 2014. Genome-wide association study of blast resistance in indica rice. BMC Plant Biol. 14, 1–11. doi:10.1186/s12870-014-0311-6

Wang, C.-H., Zheng, X.-M., Xu, Q., Yuan, X.-P., Huang, L., Zhou, H.-F., Wei, X.-H., Ge, S., 2014. Genetic diversity and classification of Oryza sativa with emphasis on Chinese rice germplasm. Heredity (Edinb). 112, 489–496. doi:10.1038/hdy.2013.130

Wang, E., Wang, J., Zhu, X., Hao, W., Wang, L., Li, Q., Zhang, L., He, W., Lu, B., Lin, H., Ma, H., Zhang, G., He, Z., 2008. Control of rice grain-filling and yield by a gene with a potential signature of domestication. Nat. Genet. 40, 1370–1374. doi:10.1038/ng.220

Wang, K., Qiu, F., Dela Paz, M.A., Zhuang, J., Xie, F., 2014. Genetic diversity and structure of improved indica rice germplasm. Plant Genet. Resour. 12, 1–7. doi:10.1017/S1479262113000579

Wang, M., Yu, Y., Haberer, G., Marri, P.R., Fan, C., Goicoechea, J.L., Zuccolo, A., Song, X., Kudrna, D., Ammiraju, J.S.S., Cossu, R.M., Maldonado, C., Chen, J., Lee, S., Sisneros, N., Baynast, K. De, Golser, W., Wissotski, M., Kim, W., Sanchez, P., Ndjiondjop, M., Sanni, K., Long, M., Carney, J., Panaud, O., Wicker, T., Machado, C.A., Chen, M., Mayer, K.F.X., Rounsley, S., Wing, R.A., 2014. The genome sequence of African rice (Oryza glaberrima) and evidence for independent domestication. Nat. Genet. 46, 982–988. doi:10.1038/ng.3044

Wang, N., Long, T., Yao, W., Xiong, L., Zhang, Q., Wu, C., 2013. Mutant Resources for the Functional Analysis of the Rice Genome. Mol. Plant 6, 596–604. doi:10.1093/mp/sss142

Wang, N., Zhang, D., Wang, Z., Xun, H., Ma, J., Wang, H., Huang, W., Liu, Y., Lin, X., Li, N., Ou, X., Zhang, C., Wang, M., Liu, B., 2014. Mutation of the RDR1 gene caused genome-wide changes in gene expression, regional variation in small RNA clusters and localized alteration in DNA methylation in rice. BMC Plant Biol. 14, 117. doi:10.1186/1471-2229-14-177

Wang, P., Kelly, S., Fouracre, J.P., Langdale, J.A., 2013. Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C4 Kranz anatomy. Plant J. 75, 656–670. doi:10.1111/tpj.12229

Wang, P., Xing, Y., Li, Z., Yu, S., 2012. Improving rice yield and quality by QTL pyramiding. Mol. Breed. 29, 903–913. doi:10.1007/s11032-011-9679-2

Wang, S., Wu, K., Yuan, Q., Liu, X., Liu, Z., Lin, X., Zeng, R., Zhu, H., Dong, G., Qian, Q., Zhang, G., Fu, X., 2012. Control of grain size, shape and quality by OsSPL16 in rice. Nat. Genet. 44, 950–954. doi:10.1038/ng.2327

Wang, T.L., Uauy, C., Robson, F., Till, B., 2012. TILLING in extremis. Plant Biotechnol J. 10, 761–772. doi:10.1111/j.1467-7652.2012.00708.x

Wang, W.Y.S., Barratt, B.J., Clayton, D.G., Todd, J.A., 2005. Genome-wide association studies: theoretical and practical concerns. Nat. Rev. Genet. 6, 109–118. doi:10.1038/nrg1522

Wang, X.-Q., Kwon, S.-W., Park, Y.-J., 2013. Evaluation of genetic diversity and linkage

disequilibrium in Korean-bred rice varieties using SSR markers. Electron. J. Biotechnol. 16. doi:10.2225/vol16-issue5-fulltext-6

Wang, Y., Li, J., 2011. Branching in rice. Curr. Opin. Plant Biol. 14, 94–99. doi:10.1016/j.pbi.2010.11.002

Wang, Y., Xiong, G., Hu, J., Jiang, L., Yu, H., Xu, J., Fang, Y., Zeng, L., Xu, E., Xu, J., Ye, W., Meng, X., Liu, R., Chen, H., Jing, Y., Wang, Y., Zhu, X., Li, J., Qian, Q., 2015. Copy number variation at the GL7 locus contributes to grain size diversity in rice. Nat. Genet. 47, 944–948. doi:10.1038/ng.3346

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63. doi:http://dx.doi.org/10.1038/nrg2484

Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. J. Amer. Static. Assoc. 58, 236–244.

Wei, X., Xu, J., Guo, H., Jiang, L., Chen, S., Yu, C., Zhou, Z., Hu, P., Zhai, H., Wan, J., 2010. DTH8 suppresses flowering in rice, influencing plant height and yield potential simultaneously. Plant Physiol. 153, 1747–1758. doi:10.1104/pp.110.156943

Weng, J., Gu, S., Wan, X., Gao, H., Guo, T., Su, N., Lei, C., Zhang, X., 2008. Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. Cell Res. 18, 1199–1209. doi:10.1038/cr.2008.307

Weng, X., Wang, L., Wang, J., Hu, Y., Du, H., Xu, C., Xing, Y., Li, X., Xiao, J., Zhang, Q., 2014. Grain number, plant height, and heading date7 is a central regulator of growth, development, and stress response. Plant Physiol 164, 735–747. doi:10.1104/pp.113.231308

Whitaker, D., Williams, E.R., John, J.A., 2001. CycDesigN: A package for the computer generation of experimental designs. CSIRO Forestry and Forest Products, CSIRO, Canberra.

Williams, W.T., 1976. Pattern Analysis in Agricultural Science. Elsevier Scientific Publishing Company, Amsterdam.

Witte, J.S., 2010. Genome-wide association studies and beyond. Annu. Rev. Public Health 31, 9–20. doi:10.1146/annurev.publhealth.012809.103723

Wu, J., Feng, F., Lian, X., Teng, X., Wei, H., Yu, H., Xie, W., Yan, M., Fan, P., Li, Y., Ma, X., Liu, H., Yu, S., Wang, G., Zhou, F., Luo, L., Mei, H., 2015. Genome-wide Association Study (GWAS) of mesocotyl elongation based on re-sequencing approach in rice. BMC Plant Biol. 15, 218. doi:10.1186/s12870-015-0608-0

Wu, K., Rao, Y., Hu, J., Zhu, G., Zhang, G., Hu, X., Guo, L., Wang, Y., Qian, Q., Zeng, D.,

2009. Characterization and fine mapping of non-panicle mutant (nop) in rice. Rice Sci. 16, 165–172. doi:10.1016/S1672-6308(08)60075-X

Xiao, J., Grandillo, S., Ahn, S.N., McCouch, S.R., Tanksley, S.D., Li, J., Yuan, L., 1996. Genes from wild rice improve yield. Nature 384, 223–224. doi:10.1038/384223a0

Xiao, J., Li, J., Grandillo, S., Ahn, S.N., Yuan, L., Tanksley, S.D., McCouch, S.R., 1998. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, Oryza rufipogon. Genetics 150, 899–909. doi:10.1016/j.tplants.2012.11.001

Xiao, J., Li, J., Yuan, L., Tanksley, S.D., 1996. Identification of QTLs affecting traits of agronomic importance in a recombinant inbred population derived from a subspecific rice cross. Theor. Appl. Genet. 92, 230–244. doi:10.1007/BF00223380

Xie, F., Guo, L., Ren, G., Hu, P., Wang, F., Xu, J., Li, X., Qiu, F., dela Paz, M.A., 2012. Genetic diversity and structure of indica rice varieties from two heterotic pools of southern China and IRRI. Plant Genet. Resour. 10, 1–8. doi:10.1017/S147926211200024X

Xie, K., Minkenberg, B., Yang, Y., 2015. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc. Natl. Acad. Sci. U. S. A. 112, 3570–3575. doi:10.1073/pnas.1420294112

Xie, X., Jin, F., Song, M.-H., Suh, J.-P., Hwang, H.-G., Kim, Y.-G., McCouch, S.R., Ahn, S.-N., 2008. Fine mapping of a yield-enhancing QTL cluster associated with transgressive variation in an Oryza sativa × O. rufipogon cross. Theor. Appl. Genet. 116, 613–622. doi:10.1007/s00122-007-0695-x

Xie, X., Song, M.-H., Jin, F., Ahn, S.-N., Suh, J.-P., Hwang, H.-G., McCouch, S.R., 2006. Fine mapping of a grain weight quantitative trait locus on rice chromosome 8 using near-isogenic lines derived from a cross between Oryza sativa and Oryza rufipogon. Theor. Appl. Genet. 113, 885–894. doi:10.1007/s00122-006-0348-5

Xing, Y., Zhang, Q., 2010. Genetic and molecular bases of rice yield. Annu. Rev. Plant Biol. 61, 421–442. doi:10.1146/annurev-arplant-042809-112209

Xing, Y.Z., Tang, W.J., Xue, W.Y., Xu, C.G., Zhang, Q., 2008. Fine mapping of a major quantitative trait loci, qSSP7, controlling the number of spikelets per panicle as a single Mendelian factor in rice. Theor. Appl. Genet. 116, 789–796. doi:10.1007/s00122-008-0711-9

Xu, J.Y., Xue, Q.Z., J., L.L., Li, Z.K., 2002. Genetic dissection of grain weight and its related traits in rice (Oryza sativa L.). Chinese J. Rice Sci. 16, 6–10.

Xu, R., Li, H., Qin, R., Wang, L., Li, L., Wei, P., Yang, J., 2014. Gene targeting using the

Agrobacterium tumefaciens-mediated CRISPR-Cas system in rice. Rice 7, 5. doi:10.1186/s12284-014-0005-6

Xu, S., Hu, Z., 2010. Methods of plant breeding in the genome era. Genet. Res. 92, 423–441. doi:10.1017/S0016672310000583

Xu, S., Yi, N., 2000. Mixed model analysis of quantitative trait loci. Proc. Natl. Acad. Sci. U. S. A. 97, 14542–14547. doi:10.1073/pnas.250235197\n250235197 [pii]

Xu, S., Zhu, D., Zhang, Q., 2014. Predicting hybrid performance in rice using genomic best linear unbiased prediction. Proc. Natl. Acad. Sci. 111, 12456–12461. doi:10.1073/pnas.1413750111

Xu, X., Liu, X., Ge, S., Jensen, J.D., Hu, F., Li, X., Dong, Y., Gutenkunst, R.N., Fang, L., Huang, L., Li, J., He, W., Zhang, G., Zheng, X., Zhang, F., Li, Y., Yu, C., Kristiansen, K., Zhang, X., Wang, J., Wright, M., McCouch, S., Nielsen, R., Wang, J., Wang, W., 2012. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat. Biotechnol. 30, 105–111. doi:10.1038/nbt.2050

Xu, Y., McCouch, S.R., Zhang, Q., 2005. How can we use genomics to improve cereals with rice as a reference genome? Plant Mol. Biol. 59, 7–26. doi:10.1007/s11103-004-4681-2

Xu, Z., Zou, F., Vision, T.J., 2005. Improving quantitative trait loci mapping resolution in experimental crosses by the use of genotypically selected samples. Genetics 170, 401–408. doi:10.1534/genetics.104.033746

Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., Zhou, H., Yu, S., Xu, C., Li, X., Zhang, Q., 2008. Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. Nat. Genet. 40, 761–767. doi:10.1038/ng.143

Yan, C.-J., Yan, S., Yang, Y.-C., Zeng, X.-H., Fang, Y.-W., Zeng, S.-Y., Tian, C.-Y., Sun, Y.-W., Tang, S.-Z., Gu, M.-H., 2009. Development of gene-tagged markers for quantitative trait loci underlying rice yield components. Euphytica 169, 215–226. doi:10.1007/s10681-009-9937-0

Yan, W., Rajcan, I., 2002. Biplot analysis of test sites and trait relations of soybean in Ontario. Crop Sci. 42, 11–20. doi:10.2135/cropsci2002.0011

Yan, W.H., Wang, P., Chen, H.X., Zhou, H.J., Li, Q.P., Wang, C.R., Ding, Z.H., Zhang, Y.S., Yu, S. Bin, Xing, Y.Z., Zhang, Q.F., 2011. A major QTL, Ghd8, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice. Mol. Plant 4, 319–330. doi:10.1093/mp/ssq070

Yang, Z., Huang, D., Tang, W., Zheng, Y., Liang, K., Cutler, A.J., Wu, W., 2013. Mapping

of Quantitative Trait Loci Underlying Cold Tolerance in Rice Seedlings via High-Throughput Sequencing of Pooled Extremes. PLoS ONE 8(7), e68433. doi:10.1371/journal.pone.0068433

Ye, G., 2010. New directions in breeding methodologies and breeding informatics, in: Board of Trustees Meeting Report, IRRI, April, 2010.

Ye, G., Collard, B.C.Y., Zhao, X.Q., Nissila, E., 2013. Enhancing rice breeding efficiency: The role of breeding informatics. Sabrao J. Breed. Genet. 45, 143–158.

Ye, G., Smith, K.F., 2010. Marker-assisted gene pyramiding for cultivar development, in: Plant Breeding Reviews. p. 33: 219–256.

Ye, G., Tu, J., Hu, C., Datta, K., Datta, S.K., 2001. Transgenic IR72 with fused Bt gene cry1Ab/cry1Ac from Bacillus thuringiensis is resistant against four lepidopteran species under field conditions. Plant Biotechnol. 18, 125–133.

Ye, G., Van Ginkel, M., 2011. Designing marker-assisted inbred line development strategies using computer simulation. Plant Breed. Rev. 34, 297–348.

Ye, G.Y., Ogbonnaya, F.C., van Ginkel, M., 2009. The use of marker- assisted recurrent backcrossing in cultivar development, in: R.K., S., Singh, R., Ye, G.Y., Selvi, A., Rao, G.P. (Eds.), Molecular Breeding in Crops, Principle, Method and Application. Studium Press:, Texas, USA:, pp. 295–319.

Ye, X., Al-Babili, S., Klöti, A., Zhang, J., Lucca, P., Beyer, P., Potrykus, I., 2000. Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 287, 303–305. doi:10.1126/science.287.5451.303

Yonemaru, J.-I., Mizobuchi, R., Kato, H., Yamamoto, T., Yamamoto, E., Matsubara, K., Hirabayashi, H., Takeuchi, Y., Tsunematsu, H., Ishii, T., Ohta, H., Maeda, H., Ebana, K., Yano, M., 2014. Genomic regions involved in yield potential detected by genome-wide association analysis in Japanese high-yielding rice cultivars. BMC Genomics 15, 346. doi:10.1186/1471-2164-15-346

Yoon, D.-B., Kang, K.-H., Kim, H.-J., Ju, H.-G., Kwon, S.-J., Suh, J.-P., Jeong, O.-Y., Ahn, S.-N., 2006. Mapping quantitative trait loci for yield components and morphological traits in an advanced backcross population between Oryza grandiglumis and the O. sativa japonica cultivar Hwaseongbyeo. Theor. Appl. Genet. 112, 1052–1062. doi:10.1007/s00122-006-0207-4

Yu, H., Xie, W., Li, J., Zhou, F., Zhang, Q., 2014. A whole-genome SNP array (RICE6K) for genomic breeding in rice. Plant Biotechnol. J. 12, 28–37. doi:10.1111/pbi.12113

Yu, J., Buckler, E.S., 2006. Genetic association mapping and genome organization of maize.

Curr. Opin. Biotechnol. 17, 155–160. doi:10.1016/j.copbio.2006.02.003

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., Buckler, E.S., 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38, 203–208. doi:10.1038/ng1702

Yu, S., Fan, Y., Yang, C., Li, X., 2008. Fine mapping of quantitative trait loci for grain length and grain width on the short arm of rice chromosome 1. Chinese J. Rice Sci. 22, 465–471 (Abstract in Chinese). doi:10.1017/CBO9781107415324.004

Yu, S., Liao, F., Wang, F., Wen, W., Li, J., Mei, H., Luo, L., 2012. Identification of Rice Transcription Factors Associated with Drought Tolerance Using the Ecotilling Method. PLoS ONE 7, e30765. doi:10.1371/journal.pone.0030765

Yu, S., Yang, C., Fan, Y., Zhuang, J., Li, X., 2008. Genetic dissection of a thousand-grain weight quantitative trait locus on rice chromosome 1. Chinese Sci. Bull. 53, 2326–2332. doi:10.1007/s11434-008-0281-x

Yu, Y., Huang, Y., Zhang, W., 2012. Changes in rice yields in China since 1980 associated with cultivar improvement, climate and crop management. Field Crops Res. 136, 65–75. doi:10.1016/j.fcr.2012.07.021

Zeng, Z.B., 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. U. S. A. 90, 10972–10976. doi:10.1073/pnas.90.23.10972

Zhang, D., Zhang, H., Wang, M., Sun, J., Qi, Y., Wang, F., Wei, X., Han, L., Wang, X., Li, Z., 2009. Genetic structure and differentiation of Oryza sativa L. in China revealed by microsatellites. Theor. Appl. Genet. 119, 1105–1117. doi:10.1007/s00122-009-1112-4

Zhang, D.L., Zhang, H.L., Qi, Y.W., Wang, M.X., Sun, J.L., Ding, L., Li, Z.C., 2013. Genetic structure and eco-geographical differentiation of cultivated Hsien rice (Oryza sativa L. subsp. indica) in China revealed by microsatellites. Chinese Sci. Bull. 58, 344–352. doi:10.1007/s11434-012-5396-4

Zhang, H., Berger, J.D., Milroy, S.P., 2013. Genotype×environment interaction studies highlight the role of phenology in specific adaptation of canola (Brassica napus) to contrasting Mediterranean climates. Field Crops Res. 144, 77–88. doi:10.1016/j.fcr.2013.01.006

Zhang, L., Li, J., Pei, Y.-F., Liu, Y., Deng, H.-W., 2009. Tests of association for quantitative traits in nuclear families using principal components to correct for population stratification. Ann. Hum. Genet. 73, 601–613. doi:10.1111/j.1469-1809.2009.00539.x

Zhang, L., Li, Q., Dong, H., He, Q., Liang, L., Tan, C., Han, Z., Yao, W., Li, G., Zhao, H., Xie, W., Xing, Y., 2015. Three CCT domain-containing genes were identified to regulate heading date by candidate gene-based association mapping and transformation in rice. Sci. Rep. 5, 7663. doi:10.1038/srep07663

Zhang, P., Li, J., Li, X., Liu, X., Zhao, X., Lu, Y., 2011. Population structure and genetic diversity in a rice core collection (Oryza sativa L.) investigated with SSR markers. PLoS ONE 6(12), e27565. doi:10.1371/journal.pone.0027565

Zhang, Q., Wang, C., Zhao, K., Zhao, Y., Caslana, V., Zhu, X., Li, D., Jiang, Q., 2001. The effectiveness of advanced rice lines with new resistance gene Xa23 to rice bacterial blight. Rice Genet. Newsl. 18, 71–72.

Zhang, Y., Luo, L., Liu, T., Xu, C., Xing, Y., 2009. Four rice QTL controlling number of spikelets per panicle expressed the characteristics of single Mendelian gene in near isogenic backgrounds. Theor. Appl. Genet. 118, 1035–44. doi:10.1007/s00122-008-0960-7

Zhang, Y., Ma, Y., Chen, Z., Zou, J., Chen, T., Li, Q., Pan, X., Zuo, S., 2015. Genome-wide association studies of international rice varieties reveal new genetic targets for improving panicle traits of tomestic rice varieties. Rice Sci. 22, 217–226. doi:10.1016/S1672-6308(14)60302-4

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., Buckler, E.S., 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42, 355–360. doi:10.1038/ng.546

Zhang, Z., Li, J., Yao, G., Zhang, H., Dou, H., Shi, H., Sun, X., Li, Z., 2011. Fine mapping and cloning of the grain umber per-panicle gene (Gnp4) on chromosome 4 in rice (Oryza sativa L.). Agric. Sci. China 10, 1825–1833. doi:10.1016/S1671-2927(11)60182-X

Zhao, K., Aranzana, M.J., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, H., Dean, C., Marjoram, P., Nordborg, M., 2007. An Arabidopsis example of association mapping in structured samples. PLoS Genet. 3(1), e4. doi:10.1371/journal.pgen.0030004

Zhao, K., Tung, C.-W., Eizenga, G.C., Wright, M.H., Ali, M.L., Price, A.H., Norton, G.J., Islam, M.R., Reynolds, A., Mezey, J., McClung, A.M., Bustamante, C.D., McCouch, S.R., 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nat. Commun. 2, 467. doi:10.1038/ncomms1467

Zhao, K., Wright, M., Kimball, J., Eizenga, G., McClung, A., Kovach, M., Tyagi, W., Ali, M.L., Tung, C.-W., Reynolds, A., Bustamante, C.D., McCouch, S.R., 2010. Genomic

diversity and introgression in O. sativa reveal the impact of domestication and breeding on the rice genome. PLoS ONE 5(5), e10780. doi:10.1371/journal.pone.0010780

Zhao, X., Zhou, L., Ponce, K., Ye, G., 2015. The usefulness of known genes/QTLs for grain quality traits in an indica population of diverse breeding lines tested using association analysis. Rice 8, 29. doi:10.1186/s12284-015-0064-3

Zhao, X., Zhou, Y., Wang, S., Xing, G., Shi, W., Xu, R., Zhu, Z., 2012. Nitrogen balance in a highly fertilized rice–wheat double-cropping system in Southern China. Soil Sci. Soc. Am. J. 76, 1068–1078. doi:10.2136/sssaj2011.0236

Zhou, H., Liu, B., Weeks, D.P., Spalding, M.H., Yang, B., 2014. Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. Nucleic Acids Res. 42, 10903–10914. doi:10.1093/nar/gku806

Zhou, J., Peng, Z., Long, J., Sosso, D., Liu, B., Eom, J., Huang, S., Liu, S., Cruz, C.V., Frommer, W.B., White, F.F., Yang, B., 2015. Gene targeting by the TAL effector PthXo2 reveals cryptic resistance gene for bacterial blight of rice. Plant J. 82, 632–643. doi:10.1111/tpj.12838

Zhou, L., Liang, S., Ponce, K., Marundon, S., Ye, G., Zhao, X., 2015. Factors affecting head rice yield and chalkiness in indica rice. Field Crops Res. 172, 1–10. doi:10.1016/j.fcr.2014.12.004

Zhu, C., Gore, M., Buckler, E.S., Yu, J., 2008. Status and prospects of association mapping in plants. Plant Genome J. 1, 5–20. doi:10.3835/plantgenome2008.02.0089

Zhu, J., Zhou, Y., Liu, Y., Wang, Z., Tang, Z., Yi, C., Tang, S., Gu, M., Liang, G., 2011. Fine mapping of a major QTL controlling panicle number in rice. Mol. Breed. 27, 171–180. doi:10.1007/s11032-010-9420-6

Zhu, M., Zhao, S., 2007. Candidate gene identification approach: progress and challenges. Int. J. Biol. Sci. 3, 420–7.

Zhu, Q., Zheng, X., Luo, J., Gaut, B.S., Ge, S., 2007. Multilocus analysis of nucleotide variation of Oryza sativa and its wild relatives: severe bottleneck during domestication of rice. Mol. Biol. Evol. 24, 875–888. doi:10.1093/molbev/msm005

Zong, G., Wang, A., Wang, L., Liang, G., Gu, M., Sang, T., Bin Han, 2012. A pyramid breeding of eight grain-yield related quantitative trait loci based on marker-assistant and phenotype selection in rice (Oryza sativa L.). J. Genet. Genomics 39, 335–350. doi:10.1016/j.jgg.2012.06.004

Zou, J., Chen, Z., Zhang, S., Zhang, W., Jiang, G., Zhao, X., Zhai, W., Pan, X., Zhu, L., 2005. Characterizations and fine mapping of a mutant gene for high tillering and dwarf in rice

(Oryza sativa L.). Planta 222, 604–612. doi:10.1007/s00425-005-0007-0

Zou, J., Zhang, S., Zhang, W., Li, G., Chen, Z., Zhai, W., Zhao, X., Pan, X., Xie, Q., Zhu, L., 2006. The rice HIGH-TILLERING DWARF1 encoding an ortholog of Arabidopsis MAX3 is required for negative regulation of the outgrowth of axillary buds. Plant J. 48, 687–698. doi:10.1111/j.1365-313X.2006.02916.x

Zuo, J., Li, J., 2014. Molecular dissection of complex agronomic traits of rice: a team effort by Chinese scientists in recent years. Natl. Sci. Rev. 1, 253–276. doi:10.1093/nsr/nwt004

Appendix

Table S1 Entry code, designation, origin, cross, basic statistics for GY of 392 rice lines across eight environments, genotype groups based on GY and subpopulations based on SSR and SNP data analysis.

Code	Designation[a]	Origin[b]	Cross[c]	Min[d]	Mean[d]	Max[d]	Sd[d]	GG[e]	Sub(SSR)[g]	Sub(SNP)[h]
1	IR73004-3-1-2-1	IRRI	IR 65469-161-2-2-3-2-2/IR 60919-150-3-3-3-2	417.2	685.2	992.4	172.8	1	2	2
2	IR 02A149	IRRI	IR00A107/PSB RC 54 (IR 60819-34-2-1)	356.5	583.9	972.7	200.1	2	1	2
3	IR 03A159	IRRI	IR 66696-49-1-2/PSB RC 64 (IR 59552-21-3-2-2)	395	594.3	1033.8	194.1	2	2	2
4	IR 05A260	IRRI	IR 72875-94-3-3-2/IR 72903-121-2-1-2	414.2	743.7	981.4	174.7	3	2	2
5	IR 07A107	IRRI	IR 64/IR 69502-6-SRN 3-UBN 1-B	414.3	679.7	941	174.5	4	2	2
6	IR 07A137	IRRI	IRRI 123/IR 77080-B-34-1-1	439.6	674	923.6	171.7	2	1	2
7	IR 09A138	IRRI	IR 75000-69-2-1-2/IR 71684-36-3-3-2	334.5	615.8	818.7	169.8	NA[f]	2	2
8	IR 08A104	IRRI	IR 72860-74-1-2-1/IR 72870-19-2-2-3	373.7	767.2	1131	272.3	5	1	2
9	IR 08A192	IRRI	NSIC RC 138/IRRI 123	576.8	733.2	950.2	124.6	6	1	1
10	IR 07A234	IRRI	NSIC RC 138/IRRI 123	365	785.9	1247.7	297.5	7	1	2
11	IR 08A128	IRRI	NSIC RC 138/IRRI 123	402.2	663.9	957.6	167.8	8	1	NA
12	IR 09A228	IRRI	PR 29232-B-17-2-1-1/IR 64	445.9	679.2	798.6	133.3	NA	2	2
13	IR 08A141	IRRI	IR 65450-173-2-1-1-3-3/IR 73000-70-2-2-2	516.1	745.6	887.5	148.2	5	1	1
14	IR 07A250	IRRI	PR 31070-4-1-2/IR 72102-4-159-1-3-3-3 (NSIC RC 112)	535.5	757.4	1105.9	184.9	6	1	2
15	IR 07A253	IRRI	PR 31070-4-1-2/IR 72102-4-159-1-3-3-3 (NSIC RC 112)	271.1	613.3	872.5	198.5	4	1	2
16	IR 07A257	IRRI	PR 31090-33-2-1/IRRI 123	440.9	761.4	1272.6	245.9	6	NA	NA
17	IR 07A260	IRRI	PR 31090-33-2-1/IRRI 123	544.1	727.9	1039.8	158.5	8	1	2
18	IR 08A138	IRRI	PR 31090-33-2-1/IRRI 123	694.7	897.7	1146.4	166.6	7	1	NA
19	IR 09A229	IRRI	PR 32749-9-10-5/PSB RC 18 (IR 51672-62-2-1-1-2-3)	592.2	874	998	148.9	NA	1	2
20	IR 09A235	IRRI	IR 77429-38-69-B-6-1-1/NSIC RC 138//IR 55423-01 (NSIC RC 9)	300	649	1042.5	256	2	1	NA

21	IR 09A220	IRRI	IR 72903-121-2-1-2/IR 71606-1-1-4-2-3-1-2 (NSIC 110	500.8	805.8	1399	281.3	7	1	NA
22	IR 09A131	IRRI	IR 77080-B-34-3/IR 71606-1-1-4-2-3-1-2 (NSIC 110	361.8	663.1	879.1	173.2	1	1	NA
23	IR 09A130	IRRI	IR 77080-B-34-3/IR 71606-1-1-4-2-3-1-2 (NSIC 110	347.5	699	1191	274.6	1	1	NA
24	IR 09A128	IRRI	IR 77776-B-8-1-2/IR 72102-4-159-1-3-3-3 (NSIC RC 112)	259.5	642	876.1	205.3	1	2	2
25	IR 09A172	IRRI	IR 73006-12-3-3-2/IR 72102-4-159-1-3-3-3 (NSIC RC 112)	267.5	591.4	810.5	165	1	1	2
26	IR 09A192	IRRI	IR 68144-2B-2-2-3-1-166/IRRI 123	407.3	728.8	1232.9	251.8	3	1	2
27	IR 09A152	IRRI	IR 84089-13/IR 72875-94-3-3-2	427.5	723.9	941.5	189.9	1	1	2
28	IR 09A181	IRRI	IR 72860-74-1-2-1/IR 73720-36-3-3-1//IR 60	505.1	706.4	928.9	162.6	6	2	2
29	IR 10A127	IRRI	PR 31070-4-1-2-PJ 26/PSB RC 64 (IR 59552-21-3-2-2)	524.6	812.3	1315.2	233.6	6	1	2
30	IR 10A128	IRRI	PR 31070-4-1-2-PJ 26/PSB RC 64 (IR 59552-21-3-2-2)	260	652.1	1108.7	318.2	NA	1	2
31	IR 10A131	IRRI	IR 77776-B-8-1-2/IR 72102-4-159-1-3-3-3 (NSIC RC 112)//IRRI 123	402.5	704	1018.4	221.8	1	2	2
32	IR06A181	IRRI	IR 71718-59-1-2-3/IR 72	457.5	765.8	1480.9	368.8	NA	2	1
33	IR 06A152	IRRI	IR01A130/JANAKI	415.6	587.8	787.6	143.3	7	2	1
34	IR 07A183	IRRI	IR 77080-B-34-1-1/IR 77776-B-8-1-2	529.4	777.8	900	115.6	4	1	2
35	IR 09A120	IRRI	JANAKI/IR02N463	566.1	877	1347.7	256.4	NA	1	NA
36	IR 09A136	IRRI	IR01A135/IRRI 123//IR01A163	523.3	717.5	940.5	152	3	1	NA
37	IR 10A125	IRRI	IR 72906-32-1-3-3/IR 72890-81-3-2	470	855.5	1315.1	258.3	NA	1	1
38	IR10A152	IRRI	IRRI 126/IRRI 164//IR05F102/IR 66946-3R-178-1-1	466	661.5	893.5	165.3	2	1	1
39	IR 10A155	IRRI	IR02W101/PSB RC 18 (IR 51672-62-2-1-1-2-3)	410	645.9	890.7	161.1	1	2	2
40	IR 10A133	IRRI	IR04A285/JANAKI	500	798.8	1075.6	201.6	9	1	2
41	IR 10A134	IRRI	IR04A285/JANAKI	475.8	755.8	913.9	150.6	5	1	2
42	IR 10A135	IRRI	IR04A285/JANAKI	485	729.7	968.2	151.1	1	1	2
43	IR10A142	IRRI	IR05A221/IR 64680-81-2-2-1-3	437.9	705.7	856.9	162.2	6	1	1
44	IR10A144	IRRI	IR05A221/IR 64680-81-2-2-1-3	367.5	705.8	1053.5	260.3	1	1	1

45	IR 71146-97-1-2-1-3	IRRI	IR00A102/IR 66452-179-2-6-1-4	335	758.8	1200.6	287.5	NA	1	1
46	IR03A477	IRRI	IR 68726-3-3-1-2/IR 71730-51-2	401.8	762.7	1171.6	243	5	2	1
47	IR04A212	IRRI	IR BB 60-1/IR 71730-51-2	389.9	772	1121.1	240.9	9	1	1
48	IR 04A216	IRRI	IR BB 60-1/IR 71730-51-2	521.2	757.8	1105.5	185.5	1	1	2
49	IR 04A285	IRRI	IR00A117/IR 64	624.9	784.6	967.2	130.3	5	1	NA
50	IR 04A395	IRRI	IR 72870-102-2-2-1-3/IR 72870-19-2-2-3	412.7	674.9	918.3	182.4	1	1	2
51	IR 09A104	IRRI	IR 66/IR 72870-19-2-2-3	462.5	727.9	909.8	160.8	9	2	NA
52	IR 10A108	IRRI	IRRI 123/IR 73000-70-2-2-2//IR 68444-18-1-3-3	468.4	780	1073	222.6	9	1	NA
53	IR 10A110	IRRI	IR05A229/IR 64	471.1	606.2	680	85	4	2	NA
54	IR 06N147	IRRI	IR02N252/IRRI 143	415	613.7	888.8	170.3	2	1	NA
55	IR 06N209	IRRI	IR 71700-247-1-1-2/PSB RC 10 (IR 50404-57-2-2-3)	395	700.7	990.8	231.9	1	1	2
56	IR 08N195	IRRI	IR 72967-12-2-3/PR 31090-33-2-1	465	707.8	940.2	160.7	1	1	1
57	IR 09N495	IRRI	IR01N200/IR01N148//IRRI 123	594.8	863.6	1233.5	234.7	NA	1	2
58	IR 09N496	IRRI	IR01N200/IR01N148//IRRI 123	357.5	932.2	1436.9	364.1	9	1	1
59	IR 09N533	IRRI	IR04N114/IR 73459-120-2-2-3//KHAO KHAE	501	650.9	781.5	88.3	4	1	2
60	IR 09N126	IRRI	IR01N111/IRRI 164	661	972.2	1472.4	272.3	3	2	2
61	IR 05N229	IRRI	IR 74052-297-2-1/IR 72165-63-2-3-3	220	740.7	1402.1	368	1	1	NA
62	IR 08N184	IRRI	IR 72967-12-2-3/IR01N142//IRRI 123	315	742.9	1440.6	343.4	7	1	2
63	IR 08N158	IRRI	IR02N444/IR 73439-11-1-3-1//IR 59682-132-1-1-2 (PSB RC 52)	482.2	698.2	900.7	155.6	8	2	2
64	IR 09N499	IRRI	IR 66/IRTP 24183	576.3	763.2	1123.5	163.2	3	2	2
65	IR 09N500	IRRI	IR 66/IRTP 24183	278.1	703.5	1275.8	310	3	2	2
66	IR 09N503	IRRI	IR 66/IRTP 24183	459.7	683	971.1	155.3	2	1	2
67	IR 09N508	IRRI	IR 69428-6-1-1-3-3/IR 64	317.5	741.6	1195.4	299.9	NA	2	2
68	IR 08N215	IRRI	IR 72875-94-3-3-2/IRTP 24183	426.8	754.5	1157.2	252.6	5	1	2
69	IR 09N509	IRRI	IR 73006-12-3-3-2/IR 72969-143-5-3-6-2	482.5	718.3	1059.5	213.5	7	1	2
70	IR 09N522	IRRI	DONGANBYEO/PSBRC 82//AREUMBYEO	307.5	797.4	1240.8	268.4	7	2	1

71	IR 09N530	IRRI	IR03N117/IR 68373-R-R-B-22-2-2//IRRI 143	422.5	694.1	1066.6	215.6	7	1	1
72	IR 09N540	IRRI	IR 55423-01 (NSIC RC 9)/IRRI 123//IR 66	372.5	752.7	1026.6	239.3	1	1	1
73	IR 09N247	IRRI	IR 68058-71-2-1/IR 71700-247-1-1-2	455	717.7	1064.2	239.4	7	1	NA
74	IR 10N134	IRRI	IRRI 143/IRTP 24183//IR 60	390.6	647.9	954.4	196.5	3	1	NA
75	IR 06N155	IRRI	IR 72158-11-5-2-3/IR 73707-45-3-2-3//IR 72875-94-3-3-2	522	703.2	850.4	115.9	2	1	2
76	IR 06N211	IRRI	IR 72158-16-3-3/BASMATI 370 (ACC 6426)	302.5	613.7	768.3	167.2	1	1	2
77	IR 08N134	IRRI	IR 72967-12-2-3/PR 31090-33-2-1	682.4	774.9	893	97.6	NA	1	2
78	IR 08N136	IRRI	IR 72967-12-2-3/PR 31090-33-2-1	329.1	823.9	1812.5	471.5	NA	2	2
79	IR 09N514	IRRI	IRRI 143/IR02N463	229.7	818.6	1340.3	339.3	5	1	2
80	IR 09N190	IRRI	DONGANBYEO/PSBRC 82//AREUMBYEO	560	884.9	1246.9	229.6	NA	1	NA
81	IR 09N527	IRRI	IR 74286-107-2-3/IR 72969-143-5-3-6-2//PSB RC 18 (IR 51672-62-2-1-1-2-3)	389.6	650.1	789.3	137	5	1	NA
82	IR 09N528	IRRI	IR 74286-107-2-3/IR 72969-143-5-3-6-2//PSB RC 18 (IR 51672-62-2-1-1-2-3)	490.6	723.3	1009.3	151.6	6	1	2
83	IR 10N272	IRRI	IR01W106/IR 71676-90-2-2	391.5	707.2	900	183.9	2	1	2
84	IR 08N138	IRRI	IR00A115/IR01N144	610.1	826.7	1031	192	NA	1	2
85	IR 10N118	IRRI	IR 68058-71-2-1/IR 71700-247-1-1-2	413.6	595.2	922.6	165.4	2	1	2
86	IR 10N108	IRRI	IR 65620-192-3-3-3-2/IR02N463//IR 72875-94-3-3-2	295	707.8	1045.6	211.8	9	1	NA
87	IR 10N186	IRRI	IR 63896-60-3-1-2/IR 71676-106-10-3//IRRI 123	495.6	738.2	1236.3	241.8	3	1	2
88	IR 10N237	IRRI	IR01N111/IRRI 164//IR 72890-81-3-2-2	500.8	716.2	993.5	159.6	2	1	NA
89	IR 10F379	IRRI	IR00A117/IRRI 149	388.8	692.9	884.1	185	2	1	1
90	IR 06M139	IRRI	MEM BERANO/PADI ABANG GOGO	527.1	690.4	820.1	124.9	NA	2	2
91	IR 06M141	IRRI	MEM BERANO/PADI ABANG GOGO	505	839.1	1749.1	401.2	3	2	2
92	IR 06M142	IRRI	MEM BERANO/PADI ABANG GOGO	495	733.4	1334.2	280.1	2	2	1
93	IR 06M150	IRRI	MEM BERANO/PADI ABANG GOGO	610	928.4	1683.6	361.3	NA	2	2
94	IR 08M113	IRRI	IR 61247-3B-8-2-1/IR 64	384.8	658.4	964.5	201.3	1	1	NA
95	IR 08M119	IRRI	IR 61247-3B-8-2-1/IRRI 123	200	695	1395.5	351.5	7	1	NA

96	IR 10M120	IRRI	IR 70114-5-3-3-3/IR 73691-14-1	235	713.9	1349	318.2	7	2	NA
97	IR 10M122	IRRI	IR 70114-5-3-3-3/IR 73691-14-1	282	786.7	1326.3	315.5	9	2	2
98	IR 10M123	IRRI	IR 70114-5-3-3-3/IR 73691-14-1	385	881.2	1284.6	284.5	7	2	NA
99	IR 02A496	IRRI	IR 72*2/W 574	430.2	663.9	970	160.7	2	1	1
100	HHZ 1-Y4-Y1	IRRI	HUANG-HUA-ZHAN*2/YUE-XIANG-ZHAN	457.4	757.8	1067.2	230.1	10	2	1
101	HHZ 12-DT 10-SAL 1-DT 1	IRRI	HUANG-HUA-ZHAN*2/TE QING	548.1	757.9	1102.1	229	3	2	1
102	HHZ 8-SAL 6-SAL 3-Y2	IRRI	HUANG-HUA-ZHAN*2/PHALGUNA	409.5	687.7	1075	208.5	10	2	1
103	HHZ 12-Y 4-DT 1-Y 1	IRRI	HUANG-HUA-ZHAN*2/TE QING	582.4	740.4	979.9	159	8	2	1
104	ZGY1	China	NA	576.8	704	869.9	113	1	2	1
105	ZH1	China	NA	570.3	799.8	1111	236.5	6	2	1
106	TME80518	IRRI	TME 80518	481.7	695.5	940	171	10	2	1
107	HHZ-12-Y4-Y3-1	IRRI	HUANG-HUA-ZHAN*2/TE QING	503.8	707.6	1158.3	207	8	2	1
108	IR 10F388	IRRI	IRRI 143/IRRI 149	412.5	671.3	839.4	131.7	2	1	1
109	HHZ5-SAL10-DT1-DT1	IRRI	HUANG-HUA-ZHAN*2/OM 1723	611.2	957.6	1677.1	346.2	3	1	1
110	IR 10F328	IRRI	IR 70215-4-CPA 3-1-3-1/IRTP 24183	435	755.9	1118.9	222.1	7	1	1
111	HHZ5-DT-8-DT1-Y1	IRRI	HUANG-HUA-ZHAN*2/OM 1723	455	741.3	1205	235.3	7	2	1
112	HHZ8-SAL9-DT2-Y1	IRRI	HUANG-HUA-ZHAN*2/PHALGUNA	348.7	733	1072.4	232.3	6	2	1
113	IR 10F339	IRRI	IR 70215-4-CPA 3-1-3-1/IRTP 24183	475	678.3	938.5	167.8	2	1	1
114	IR 04A115	IRRI	IR 68077-82-2-2-2-3/IR00A117	347.5	877.6	1608.1	401.6	7	1	NA
115	IR 03A290	IRRI	IR 68068-99-1-3-3-3/JANAKI//PSB RC 18 (IR 51672-62-2-1-1-2-3)	325	706	1420.1	360.7	7	1	2
116	IR 04A409	IRRI	IR00A112/PSB RC 20 (IR 57301-195-3-3)	250	646.2	841.3	197	1	2	2
117	IR06A145	IRRI	IR02A127/JANAKI	510	662.5	820.1	135.4	NA	2	1
118	IR 06A150	IRRI	IR02A127/IR 64	622.1	797.8	916.2	121	NA	2	NA
119	IR 07A179	IRRI	IR01A135/IR 77080-B-34-3	506	735.6	1089.3	182.3	3	1	1
120	IR 08A175	IRRI	IR 73013-95-1-3-2/IRRI 123	470.4	655.3	838.8	135.8	5	1	2
121	08A191	IRRI	IR02A149/IRRI 123	372.5	733.2	1174.4	228.3	7	1	NA

122	IR 09A116	IRRI	IR BB 60-1/IRTP 24183	412.5	654.9	889.7	164.5	7	1	2
123	IR 09A231	IRRI	IR 73013-95-1-3-2/IRRI 123//IR01A163	365.9	611.8	775	136.6	NA	1	2
124	IR 07A166	IRRI	IR 73013-95-1-3-2/IR 72862-27-3-2-3	350	578.8	764.2	140.3	2	1	NA
125	IR 09N272	IRRI	IR 74286-107-2-3/IR 72969-143-5-3-6-2//PSB RC 18 (IR 51672-62-2-1-1-2-3)	464.1	661.4	829.5	130.7	NA	2	1
126	IR 06N233	IRRI	IR 72158-16-3-3/IR01A154//PSB RC 64 (IR 59552-21-3-2-2)	541.9	885.1	1745.6	435.6	NA	2	2
127	IR 10N211	IRRI	IR 72967-12-2-3/IR 72164-348-6-2-2-2//IR 71606-1-1-4-2-3-1-2 (NSIC 110	398.5	605.9	748.8	123	8	1	2
128	IR 10N251	IRRI	IR BB 60-1/IR 73711-130-1-3-1//IR 65450-173-2-1-1-3-3	431	589.6	961.2	178.3	NA	2	2
129	IR 10N225	IRRI	IR01N106/IR01N194//IRRI 143	435	599.2	953.2	211.7	NA	2	2
130	IR 10N226	IRRI	IR01N106/IR01N194//IRRI 143	277	693	977.7	243.1	NA	1	NA
131	IR 10N230	IRRI	IR01N106/IR01N194//IRRI 143	478.8	781.9	1431.1	315.2	3	2	2
132	IR 10N291	IRRI	IR05N341/IR 64680-81-2-2-1-3	379	715.9	1059.8	218.9	1	2	2
133	IR 10N303	IRRI	IR 72906-32-1-3-3/IR02N139//PSB RC 64 (IR 59552-21-3-2-2)	592.5	862.8	1315.6	314.4	NA	2	2
134	IR 10N304	IRRI	IR 72906-32-1-3-3/IR02N139//PSB RC 64 (IR 59552-21-3-2-2)	624.5	835.4	1243	206.6	9	2	2
135	IR 05N168	IRRI	IR 72889-69-2-2-2/IR 72158-68-6-3	330	606.5	937.7	211	7	2	NA
136	IR 05N170	IRRI	IR 72889-69-2-2-2/IR 72158-68-6-3	365	746.9	1090.8	235	NA	1	2
137	IR 06N234	IRRI	IR 72158-16-3-3/IR01A154//PSB RC 64 (IR 59552-21-3-2-2)	428.9	657	845.9	134.6	8	1	2
138	IR 07N123	IRRI	IRRI 164/IR 68058-71-2-1//IR 72890-81-3-2-2	427.4	659.1	812.4	156	1	2	1
139	IR 09N146	IRRI	IR 73012-15-2-2-1/IR01N144	512.8	807.4	1311.9	255.4	3	1	1
140	IR 09N142	IRRI	IR 73012-15-2-2-1/IR01N144	421.7	618.9	953.1	181.7	2	1	2
141	IR 10F403	IRRI	IR 69726-29-1-2-2-2 (MATATAG 2)/IRRI 149	376.6	695.7	920	164.1	5	2	1
142	IR 10F202	IRRI	IR 80410-B-197-4/IRRI 149	485.6	724.1	1148	290.8	NA	2	1
143	IR 10F371	IRRI	IR 70215-4-CPA 3-1-3-1/IRTP 24183	384.6	589	761.2	112	4	1	1

144	IR 10F203	IRRI	IR 80410-B-197-4/IRRI 149	480.5	737.7	1020.8	194.9	5	1	1
145	SACG-4	China	NA	381.8	683.5	955.1	213.8	NA	NA	NA
146	C7546WH-2-2-1-1-4-2-1-2-1-1-2	PhilRice	PR 30248 WH 18-C2-3-1-1//JHR 3 B-CB-8/IR 68284 H-C54-2-1	385	552.4	652.2	82.8	1	2	1
147	PR30245-10-414	PhilRice	IR64-AC97 WP (ANTHER CULTURE)/?	399.8	621.7	794.9	136.7	4	1	1
148	PR34859-B-4-1-1-2-1(G)	PhilRice	MALIGAYA SPECIAL # 6/GUANDONG 7	330	607.3	856.4	171.9	2	2	1
149	PR35251-2B-5-5-3-1-1	PhilRice	YANG YU/FR 4	503.6	756.2	946.3	170.8	9	2	1
150	MTS-1681	IRRI	NA	344.1	623.6	917.4	195.8	2	1	1
151	PR35769-B-37-3-1-2-1	PhilRice	PSB RC 28/IR 65907-116-1-B	286.5	738.5	1235.1	282.8	7	1	1
152	PR35786-B-3-1-3-1-1-3	PhilRice	AR 32-19-3-3/PR 33214	495	650.8	841.9	134.4	NA	1	1
153	PR35786-B-3-3-2-1-1	PhilRice	AR 32-19-3-3/PR 33214	355.6	708.6	924.1	188.8	2	2	1
154	PR35805-B-9-2-3-2-3	PhilRice	BURDAGOL (A)/IR 74286-55-2-3-2-3	476.5	670.4	812.6	102.2	5	2	1
155	PR36723-B-1-3-3-3-2	PhilRice	ANGELICA/IIRON 327 A 4 WS	481.9	729	1107	207.6	7	2	1
156	PR37139-3-1-3-1-2-1	PhilRice	ANGELICA/PULANG HUMOT//AR 32-19-3-3	388.2	703.3	870.5	157.8	4	2	1
157	PR37152-2-2-4-1-1-1	PhilRice	AR 32-4-5-2//PR 34056-B-3-1-1/BURDAGOL	530	712.8	901	154.8	NA	2	1
158	PR37171-1-1-1-2-1-1-1	PhilRice	AR 32-19-3-4/MATATAG 6	282.1	561.6	779	172.1	4	2	1
159	PR37246-2-3-2-1-1-2-1	PhilRice	IR 60819 R/AR 32-19-3-3	524.2	730.2	838.8	134.9	2	2	1
160	PR37252-2-1-1-1-2-2	PhilRice	RIL 467/AR 32-4-58-2	479.5	792.1	1251	243	NA	1	1
161	PR37704-2B-6-1-2-1-1	PhilRice	PR 33212-CB-1/C 6537-56-1	523.8	869	1785.2	393.7	3	2	1
162	PR37921-B-3-4-2-1-2	PhilRice	ADRON 117//DM 25/AR 32-19-3-3	484.2	664.2	761.6	96.5	NA	2	1
163	IR08N113	IRRI	PSB RC 64 (IR 59552-21-3-2-2)/KHAO' LERN//IR 74286-107-2-3	330	676.1	1027.7	209	NA	1	1
164	PR37942-3B-5-3-2	PhilRice	PR 35796-B-3-2-3/ADRON 111	414.4	714.8	893.2	178.6	8	1	1
165	PR37951-3B-37-1	PhilRice	PR 36828-38-1-1/PR35786-B-3-1-4	507.2	785.5	1005.4	187.8	5	1	1
166	PR37951-3B-37-1-2	PhilRice	PR 36828-38-1-1/PR35786-B-3-1-4	473.6	792.7	1159.4	285.7	10	2	1
167	PR37952-B-1-1-2	PhilRice	PR 29740-B-26-1-2-1-1-1-1/C 6537-56-1	513.9	932.5	1554.1	359.5	3	2	1
168	IR10N293	IRRI	IR05N341/IR 64680-81-2-2-1-3	533.2	715.3	983.5	171.7	1	NA	NA

169	PR37246-2-3-2-1-1-2-2	PhilRice	IR 60819 R/AR 32-19-3-3	430	656.5	888.8	175.2	2	2	1
170	PR37990-3B-15-2	PhilRice	PR 36925-B-2/PR 33212-CB-1	473.8	741.8	997.5	180.3	NA	1	1
171	PR38012-3B-3-1	PhilRice	PR 34544-B-7-3-2-1-1-2-1//C 6688 WH6-2-2-1/PR 31563	520.2	873.5	1660.7	385.4	3	1	1
172	PR37952-B-4-1-3	PhilRice	PR 29740-B-26-1-2-1-1-1-1-1/C 6537-56-1	414.8	682.4	897.2	179.5	7	1	1
173	PR40078-B-12-2	PhilRice	NA	507.2	762.2	929.6	133	6	2	1
174	PR40083-1B-3-2	PhilRice	NA	381.3	624.4	764.8	129.1	4	1	1
175	IR 09N542	IRRI	IR 69726-116-1-3 (MATATAG 1)/IR 72967-170-1-2-3-2	460.8	697.6	877.8	146.2	2	NA	NA
176	IR 78222-20-7-148-2-B	IRRI	IR 75499-14-1-B/IR 75499-6-1-B	176.7	463.7	852.4	226.5	NA	2	1
177	IR 77186-148-3-4-3	IRRI	IR 73885-1-4-3-2-1-6 (MATATAG 9)/IR 70479-45-2-3//IR 64680-81-2-2-1-3	562.7	777.8	1095.8	180.3	NA	1	1
178	HHZ 12-DT-10-SAL-1-DT1-b	IRRI	HUANG-HUA-ZHAN*2/TE QING	505.3	771.8	1301.7	260.8	3	2	2
179	BR29	Bangladesh	BG90-2/BR 10 (BR 51-46-5)	485	859	1065.6	188.8	9	2	1
180	Ciherang	IRRI	IR 18349-53-1-3-1-3/2*IR 19661-131-3-1-3//IR 5657-33-2-1/IR 2061-465-1-5-5	319.5	721.1	1111	242	3	2	1
181	MTU1115	India	SAMBA MAHSURI/DP 13	272.7	578.2	895.1	195	8	2	1
182	YTL126	IRRI	NA	463.6	656.3	872	123.4	3	1	1
183	NSICRc 214	IRRI	IR00A112/PSB RC 20 (IR 57301-195-3-3)	461.3	661.6	812	136.1	NA	1	1
184	as996	IRRI	IR 64/O RUFIPOGON	586	744.1	945.1	130.3	5	2	1
185	YD6	China	Yangdao4/Yan3021	521.8	721.1	980.8	160.3	8	2	1
186	IR 87520-44-3-1-2	IRRI	IR 72967-12-2-3/IR BB 60-1	492.5	699.8	889.7	141.7	7	2	1
187	IR66BB	IRRI	NA	395.7	663.3	804.7	149.3	5	2	1
188	IR 71033-121-15	IRRI	IR 31917-45-3-2*4/W 1342	368	735	976.4	195.9	4	2	1
189	YUNJING 23	China	NA	272.5	526.9	813.7	177.2	NA	NA	NA
190	PSBRC86	IRRI	IR 10198-66-2/TCCP 266-B-B-B-10-3-1	300.4	587.4	775	142.9	4	NA	NA
191	IR 03A568	IRRI	IRRI 145/PSB RC 18 (IR 51672-62-2-1-1-2-3)//IRRI 145	301.8	614	900	204.2	4	2	NA
192	OM 6378	Vietnam	TYPE 3/JASMINE 85	207.9	661.5	1064.2	283.6	4	2	1
193	GHAIYA	Nepal	MTU 15 (ACC 233)/WAIKOKU	456.1	680.2	928	168.4	NA	NA	NA

194	IR 80340-12-B-B-1-2-B-B	IRRI	NA	530.5	775.5	954	142.4	NA	2	1
195	SAGC-08	China	NA	339.2	574.3	930.1	210.9	NA	NA	NA
196	IR 06A145-B	IRRI	IR02A127/JANAKI	477.7	778.8	1076.2	196.7	5	2	1
197	OM 6073	Vietnam	C 3/D 3//D 3	258.7	672.3	977.5	234.8	10	1	1
198	IR 58	IRRI	IR 28/KWANG CHANG AI//IR 36	209.1	532.5	860.3	245.8	6	2	1
199	CT 15679-17-1-1-2-3-M	CIAT	ECIA 213-F 4-J 153/RHS 376-57-CX-2 CX-3 CX OZA//CPR 32	281.7	647.2	1050	236.1	10	2	1
200	IR 06M141-B	IRRI	MEM BERANO/PADI ABANG GOGO	566.7	730.7	840	96	NA	NA	NA
201	IR 68	IRRI	IR 19660-73-4/IR 2415-90-4-3-2//IR 54	435.9	690.6	997.5	170.4	10	2	1
202	IR 78545-49-2-2-2	IRRI	IR 71137-184-3-2-3-3/IR 72875-94-3-3-2	353.8	661	860.9	162	8	1	1
203	IR 56	IRRI	IR 4432-53-33/PTB 33//IR 36	398.3	587.4	830	159.8	NA	2	1
204	PADI LIWAGU	Indonesia	NA	274.3	507.4	1020	229.2	NA	NA	NA
205	IR 04A421	IRRI	IR 73012-137-2-2-2/PSB RC 10 (IR 50404-57-2-2-3)	611.6	772.2	933.6	121.8	8	1	1
206	BP 10618F-BB8-13-BB8	Indonesia	CIHERANG/IR BB 11	453	663.3	1009.4	175.3	2	1	1
207	IR 36	IRRI	IR 1561-228-1-2/IR 1737//CR 94-13	357.4	565.8	872	180.8	8	2	1
208	HHZ 5-SAL 9-Y 3-Y 1	IRRI	HUANG-HUA-ZHAN*2/OM 1723	411.7	688.8	1160.6	232.2	8	2	1
209	OM 6600	Vietnam	D 43/JASMINE 85	530.6	749.8	1008.1	161.4	6	2	1
210	IR 08M110	IRRI	IR 75493-8-2-1-2-3/IR 65620-192-3-3-3-2	430.5	706.3	945	165.6	4	2	1
211	IR 07A179-b	IRRI	IR01A135/IR 77080-B-34-3	345.4	618.8	896.6	191.9	6	2	2
212	IR 02A201	IRRI	IR 73887-1-8-3-5/IR00A107	403.5	606.3	915	152.3	8	1	1
213	IR 80340-23-B-12-6-B	IRRI	NA	379.4	598.2	916.1	169.7	NA	2	1
214	NSICRC222	IRRI	IR 73012-137-2-2-2/PSB RC 10 (IR 50404-57-2-2-3)	441.7	719.7	1032.7	217.5	8	2	1
215	IR 01A160	IRRI	IR00A107/IR 65620-192-3-3-3-2	425	639.9	909.2	174.8	2	2	1
216	IR 72892-77-2-2-2	IRRI	IR 61961-159-2-3-3-2-2/IR 65629-67-3-3-1-1-2	419.7	707.8	900	170.7	8	1	1
217	IR 06M145	IRRI	MEM BERANO/PADI ABANG GOGO	517.6	682.6	829.2	113.9	4	1	2
218	OM 5628	Vietnam	IR 64/2*C 54	351.5	673.2	1017.5	227.5	4	1	1
219	IR 82870-57	IRRI	IR 82851-16/IR 82855-9	383.7	619.2	781.5	152.5	4	1	1

220	PR 35887-1-21-2-1	PhilRice	ISH 58/MATATAG 6	325	623.9	875	210.3	1	2	1
221	PR 33319-9-1-1-5-3-5-4-1	PhilRice	PR 26871-27-3-PJ 17/PR 26850-PJ 18-13-8	347.6	583.1	672.2	102.2	4	2	1
222	IRRI 139	IRRI	IR 72/ZAWA BONDAY	213.8	475.5	620.2	139.2	4	2	1
223	PSBRC84	IRRI	CSR 10/TCCP 266-B-B-B-10-3-1	270.7	562.7	796.9	164.8	4	2	1
224	GZ 7712-BC-H-6-20	Egypt	SAKHA 101*2/GZ 5310-20	155	497.1	865	273.5	NA	NA	NA
225	IR-BB57(IR72919-10-1-3)	IRRI	AY 4+5/IR 66700-4-2-9-5-2	435.1	658.5	932.5	172.4	8	2	1
226	IRRI 129	IRRI	IR 2053-521-1-1-1/K 116//KN-1B-361-1-8-6-9-1	331.8	572.8	859.7	172.8	4	2	1
227	IR 73887-1-8-1-4	IRRI	IR 84685/IR 64	441	716	1011.1	177.1	5	NA	NA
228	IR 28	IRRI	IR 833-6-2-1-1//IR 1561-149-1/IR 1737	316	555.1	760	159	10	1	1
229	PR 37264-1-4-1-1-3	PhilRice	PR 31561-AR 32-11-76-9-2-1-B-B-MB-1/PR 31560-AR 32-2-66-4-3-1-B-B-MB-1	439.6	657.8	941.5	187.2	1	2	1
230	IR 02A127	IRRI	IR00A107/IR 62243-41-1-3-3	495	713	893.5	128.4	NA	2	1
231	PR 33282-B-8-1-1-1-1-1	PhilRice	ZHANGYU 87/PJG 6	365	641.5	882.5	179.9	1	2	1
232	OMCS 2000	Vietnam	OM 1738/MRC 19399	264.1	639.8	983.6	261.4	8	1	1
233	CT 15716-6-1-2-2-2-M	CIAT	UQUIHUA/JUMA 62//CT 9748-13-2-1-M-M-1-1	325	663.5	1173	268	NA	2	1
234	CT 18148-6-9-5-1-2-MMP	CIAT	CT 9748-3-1-1P-2-M/CAPIRONA//BCF1720	370.8	611.4	883.9	177.9	1	2	1
235	WAS 161-B-6-4-FKR 1 (NERICA-L-35)	AfricaRice	WAB 1291/4*IR 64	362.6	588.7	1012.5	211.2	NA	1	1
236	CT 15673-8-1-4-1-6-M	CIAT	CT 7948-AM-14-3-1/C 109 CU 84//CNAX 5011-9-1-6-4-B	419.7	561.8	773.4	148.2	NA	2	1
237	IR 78875-190-B-1-3	IRRI	IR 55423-01 (NSIC RC 9)/IR 64	190.6	559.9	847.5	243.5	4	2	1
238	PSBRC60	IRRI	IR 8234-OT-9-2/IR 19661-131-1-2	440	626.9	900.5	155.4	1	NA	NA
239	SAGC-02	China	NA	561.5	717.5	979	141.5	NA	2	1
240	IR 05N173	IRRI	IR 72889-69-2-2-2/IR 72158-68-6-3	316.2	750.5	1304.7	286.3	3	2	2
241	IR 74095-AC45	IRRI	M 202/OU 301	268	596.2	867.5	179.8	4	2	1
242	IR 50	IRRI	IR 2153-14-1-6-2/IR 28//IR 36	302.3	563.9	770	181.8	6	2	1
243	B11143D-MR-1-PN-3-MR-3-SI-2-	Indonesia	BENORAJA/FATMAWATI//TUKAD UNDA	450.5	609.4	714.8	79.3	4	2	1

	3-PN-1									
244	MILYANG 46(CHEONGCHEONGBYEO)	Republic of Korea	YR 675-153-2-2/IR 2035-290-2	375.4	605.2	975.9	211.8	NA	NA	NA
245	IR 05N412	IRRI	IR 72875-94-3-3-2/IR 73707-45-3-2-3	380	701.6	979.3	200.2	2	1	2
246	IR 71033-4-1-127-B	IRRI	IR 31917-45-3-2*4/W 1342	321.5	627.9	800	172.6	8	2	1
247	IR 77384-12-35-3-12-1-B	IRRI	NA	168.8	426	543.5	123.5	4	2	1
248	IR 03A262	IRRI	IR 71606-1-2-1-3-2-3-1/PSB RC 64 (IR 59552-21-3-2-2)	497	730.8	842.2	138.1	8	1	2
249	IR 06N234-b	IRRI	IR 72158-16-3-3/IR01A154//PSB RC 64 (IR 59552-21-3-2-2)	406.9	590.9	756.9	113.6	NA	1	2
250	BR 7414-22-1	Bangladesh	NA	431.5	759.9	1120	213.7	10	2	1
251	2001059-TR 2151-6-1-1	Turkey	ZENA/KIRAL	123.7	407	712.5	196.8	NA	NA	NA
252	WAB96-1-1	AfricaRice	ITA 257/YS 121	403.8	583.2	753.8	125.6	1	2	1
253	CT 18685-10-1-1-3-3	CIAT	CT 18069-24(1)/EPAGRI 108	460	687	814.9	123.8	8	2	1
254	IR 55423-01	IRRI	UPL RI 5/IR 12979-24-1 (BROWN)	320	688.2	950	211.6	4	2	1
255	IR 05A278	IRRI	IR 73003-151-2-3-1/IR00A115	535.3	674.8	815.8	116.3	8	2	1
256	IR 07T102	IRRI	IR 69726-80-1-3/IR 64680-81-2-2-1-3	447.1	689.7	896.4	163.1	NA	2	1
257	MATATAG2	IRRI	IR 52256-84-2-3/IR 72//IR 1561-228-3*2/UTRI MERAH (ACC 16682)	384.6	580.2	689.3	110.8	3	1	1
258	IR 65483-111-5-9-2-11	IRRI	IR 56*4/O BRACHYANTHA	231	493.1	930	214.3	10	2	1
259	BR 7232-6-2-3	Bangladesh	NA	229.1	650.7	970	224.8	4	2	1
260	PK 7909-3-1-2-2	Pakistan	70007/GM SUPER	292.7	587.7	879.6	169.6	8	1	1
261	CT 15671-15-4-5-1-1-M	CIAT	RHS 376-57-CX-2 CX-3 CX OZA/PUSA 169//CT 11275-3-F4-8P-2	414.1	647.2	915	166.9	10	2	1
262	IR 07A253-b	IRRI	PR 31070-4-1-2/IR 72102-4-159-1-3-3-3 (NSIC RC 112)	371.5	700.7	1070	223.2	10	1	2
263	IR-BB56(IR72918-37-1-1)	IRRI	AY 4+5/IR 68311-13-3-42	456.7	651.4	1000	203.3	10	2	1
264	PSBRC82	IRRI	IR 47761-27-1-3-6/PSB RC 28 (IR 56381-139-2-2)	471	684	777.9	105.9	4	1	1
265	PSB RC 82-SUB1	IRRI	IRRI 123*2/IRRI 149	315.9	607.3	817.7	163.2	4	1	1

266	HUANGHUAZHAN	China	HUANGXINZHAN/FENGHUAZHAN	551.2	790.3	1087.5	195.5	10	2	1
267	TE QING	China	YE QING LUN/TE AI	480.3	714.1	1097.5	207.3	10	2	1
268	NSIC RC152	IRRI	PR 26134-12-6-3-1/PR 26684-31-2-1-5-1	271.3	675	937.5	202.7	4	2	1
269	IR 71146-97-1-2-1-3-b	IRRI	IR00A102/IR 66452-179-2-6-1-4	347.5	632	1064.9	216.2	8	1	2
270	RADHA 4	Nepal	BG 34-8/IR 36	505.6	719	1020	170.5	NA	NA	NA
271	IR 43	IRRI	IR 305-3-17-1-3/IR 24	52.9	393.3	905.3	349.1	NA	NA	NA
272	ZHONGZU 14	China	Wufengzhan2/IRBB5//Wufengzhan2	406.7	699.6	1132.5	227.3	10	2	1
273	WAS 122-IDSA-1-WAS-2-WAB 2-TGR 7 (NERICA-L-17)	AfricaRice	WAB 1291/3*IR 64	434.5	586	728.3	108.5	5	2	1
274	IR 72	IRRI	IR 19661-9-2-3-3/IR 15795-199-3-3//IR 9129-209-2-2-2-1	413.1	665.8	935	170.2	10	2	1
275	IR 06M143	IRRI	MEM BERANO/PADI ABANG GOGO	561.2	750.1	907.5	121.8	8	2	2
276	IR 06N146	IRRI	IR01N200/IR 1529-ECIA	525.1	827.7	1380	272.9	10	1	2
277	PAU-201	India	BJ 1/IR 6-156	487.8	624	832.5	128.7	NA	2	1
278	FFZ 1	China	NA	382.7	641.5	1018	203	2	2	1
279	IR 08A131	IRRI	NSIC RC 138/IRRI 123	447	704.4	911.3	194.5	1	1	1
280	BP 1356-1G-KN-4	Indonesia	TAJUM*2/MAROS	361.8	701.1	1192.5	252.6	10	1	1
281	A69-1	Sri Lanka	BG 94-1/POKKALI	446.2	682.7	926.2	158.4	NA	NA	NA
282	IR 04N106	IRRI	IR 68544-29-2-1-3-1-2/IR 73887-1-8-2-1//IR 59682-132-1-1-2 (PSB RC 52)	418.7	731.3	1060	208.9	10	2	1
283	IRGA 318-11-6-9-2B	IRRI	NEWREX//IR 19743-25-2-2/BR IRGA 409	474.7	622.8	875	124.1	10	2	1
284	CT 18154-5-1-4-2-2-M	CIAT	CT 8455-1-24-1P-1X/CAPIRONA//BCF1720	426.3	593.7	754.9	108.7	3	2	1
285	IR 79233-1-2-1-2	IRRI	IR 72906-32-1-3-3/IR 72158-16-3-3	524.7	803.4	1297.5	260.9	10	2	1
286	IR 24	IRRI	IR 8/IR 127-2-2	415.3	679.1	1070	200.2	10	2	1
287	ADRON 125	Suriname	RCN B-93-216/WAB 450-I-B-P-163-4-1	424.1	604.2	1015	189.7	NA	NA	NA
288	CT 15765-13-3-6-2-1-M	CIAT	CT 9509-17-7-1P-1PT//CT 5747-38-1-1-1P/COSTA NORTE	493.8	730.7	1105	195.6	10	2	1
289	IR 44	IRRI	IR 1529/CR 94-13//IR 480-5-9-3	355.7	612.3	892.2	190.5	2	2	1

290	NR 11	Vietnam	424/CR 203	464.1	704.2	1060	205.7	10	2	1
291	CT 19561-3-20-2-3-2-2-M	CIAT	CT 18487-F2//POOLBCF/PI9	343.6	612.6	869.6	176.8	6	2	1
292	IR 78555-68-3-3-3	IRRI	IR 72870-102-2-2-1-3/IR 72870-19-2-2-3	466	661.3	943.7	188.5	8	2	1
293	IR 06N155-B	IRRI	IR 72158-11-5-2-3/IR 73707-45-3-2-3//IR 72875-94-3-3-2	468.1	734	1055	197.4	10	1	2
294	IR 04A428	IRRI	IR 73718-1-2-1-3/PSB RC 10 (IR 50404-57-2-2-3)	515.6	820.2	1149.7	216.6	NA	2	2
295	IR 60	IRRI	IR 4432-53-33/PTB 33//IR 36	410.8	591.4	960	193.7	10	2	1
296	PK 7392-10-1-1-1-1	Pakistan	4048/ 5005-4	247.8	540.9	957.8	225.6	1	2	1
297	IR 07T104	IRRI	IR 71606-1-2-1-3-2-3-1/IR 55182-2B-14-3-2	506.8	716.4	864.6	131.3	5	1	1
298	C 3419-10-1-2 (PSB RC98)	PhilRice	IR 4563-52-1-3-6/IR 36974-13-3-3//IR 50404	486.7	747.9	960.9	158.6	8	2	1
299	HARDINATH 1	Nepal	BG 951/// 79-3348/H 4//BW 288-1-3	429.5	675.7	960	179.8	NA	NA	NA
300	IR 03A550	IRRI	IR 68427-15-2-3-1/IR 68068-99-1-3-3-3//IR 1529-ECIA	446.7	778.4	1036.9	206.9	8	2	NA
301	KHUDWANI ACC 193	India	NA	183.2	419.6	731.9	175.8	NA	NA	NA
302	IR 01A122	IRRI	IR 58773-35-3-1-2/IR 65441-170-2-3-3-2-3//IR 64	401	688.2	980	181.1	10	1	1
303	IR 03A500	IRRI	IR 69713-43-1-3-2-3/IR 59682-132-1-1-2 (PSB RC 52)	418.6	615.4	828.9	166.1	NA	2	2
304	CT 17379-32-5-1-1-5-M	CIAT	CT 14938-36-1-M/CT 10532-1-1-2-1-1 T-3 P//CT 10532-1-1-2-1-1 T-3 P	385	562.7	939.8	193.3	8	2	1
305	IR 05A272	IRRI	IR 72904-65-1-3-3/IR 73012-137-2-2-2	527.7	663	776.8	102.9	NA	1	1
306	IR64	IRRI	IR 5657-33-2-1/IR 2061-465-1-5-5	500.1	687.2	1022.5	156.5	10	1	1
307	OM 6610	Vietnam	AS 996/OM 4900	436.8	717.9	1008	210.5	10	2	1
308	IR 10T113	IRRI	IR 68144-2B-2-2-3-1/IR 66946-3R-78-1-1//IR 77080-B-4-2-2	413.1	640.5	1030	226	10	1	1
309	IR 05N386	IRRI	IR 72985-65-3-1/IR 72903-121-2-1-2	379.3	637.6	885.7	164.3	3	1	1
310	CT 15675-7-1-7-1-2-M	CIAT	CT 7948-AM-14-3-1/C 109 CU 84//SELVA ALTA	617.8	799.5	1230.6	206.5	3	2	1
311	6527	China	Lvhan1 mutant	425	629.8	1100	212.9	10	2	1
312	MALA	Bangladesh	B 5580 A 1-15*2/SIGADIS	204.2	529.9	914	230.1	8	2	1
313	IR 05A272-B	IRRI	IR 72904-65-1-3-3/IR 73012-137-2-2-2	503.6	706.6	1065	179.8	10	1	2
314	CT 15691-4-3-3-1-1-M	CIAT	ECIA 213-F 4-J 153/CT 6142-F 2-RH-3-4-3//EPAGRI 108	399.9	674.7	952.5	199.5	10	2	1

315	PR-114	PhilRice	TN 1/PATONG 32//PR 106*4///IR 8	373.2	596.7	903.3	173.6	2	2	1
316	WEED TOLERANT RICE 1	China	WEED TOLERANT RICE 1	417.4	722.2	1200	281.8	NA	NA	NA
317	IR 83265-1-1-13-40-1-10-1-1-1-1	IRRI	MANGEUMBYEO/IR BB 57 (IR 72919-10-1-3)	111.1	481.1	1010	315.9	NA	NA	NA
318	IR 66	IRRI	IR 13240-108-2-2-3/IR 9129-209-2-2-2-1	391.1	617.4	713.5	105.6	4	2	1
319	FEDEARROZ 50	Colombia	NA	463.1	688.4	1167.9	215	NA	NA	NA
320	BP 10620F-BB8-15-BB4	Indonesia	CIHERANG/IR BB 64	426.9	673.5	881.2	154.6	8	2	1
321	WANXIAN 763	China	NA	439.5	642.7	944.5	194.7	8	2	1
322	SONALEE (HIGH PROTEIN)	India	NA	401.1	574.7	705.2	107	6	2	1
323	IR 68552-55-3-2	IRRI	IR 66159-52-2/GUNDIL KUNING	420	511.1	650.7	88	NA	NA	NA
324	ZX115	China	NA	300.8	612.1	1148.3	257.6	8	2	1
325	NSICRC212	IRRI	IR 68077-82-2-2-2-3/IR 59548-122-1-4-1	490.8	764.3	1019.8	197.5	6	1	1
326	IR 08A176	IRRI	IR 73013-95-1-3-2/IRRI 123	525.3	706.9	910.1	120.2	8	2	NA
327	CT 15672-12-1-5-2-4-M	CIAT	CT 7948-AM-14-3-1/C 109 CU 84//CT 7948-8-4-1 P-2 X	433.6	575.6	669.5	82.5	NA	2	1
328	99035-TR 2002-2-2-1	Turkey	IR 66160-5-2-3-2/VENERIA	138.2	306.5	618.6	188.8	NA	NA	NA
329	IR 74099-AC 7	IRRI	DAEYABYEO/GIZA 177	186.4	562	719.2	186.2	4	2	1
330	IR 05N496	IRRI	IR 74052-297-2-1/IR 71700-247-1-1-2//IR 73885-1-4-3-2-1-6 (MATATAG 9)	399.3	657.1	845.1	149.4	4	1	2
331	YUNDAO 1	China	IRGC10203/Boro5//Dianxi1///Hongza135	341.6	520	820.4	164.8	NA	NA	NA
332	IR 78554-145-1-3-2	IRRI	IR 72861-13-2-1-2/IR 68450-36-3-2-2-3	357.4	610.4	865.3	149.2	6	1	1
333	OM 4900	Vietnam	C 53/JASMINE 85//JASMINE 85	369.7	700.1	1026.7	185.6	8	2	1
334	PSBRC96	IRRI	IR 50404//DOBONGBYEO/MOROBEREKAN	338	601.2	885	159.9	10	2	1
335	IR 07L167	IRRI	IR 57514-PMI 5-B-1-2/IRRI 123	288.9	654.9	850.8	181.4	4	1	1
336	IR 81352-65-2-1-3-3	IRRI	IR 73013-95-1-3-2/IR 72862-27-3-2-3	405	607.8	755	135.1	NA	2	1
337	OMCS 2009	Vietnam	OM 1314/OM 2514	375	625.9	779.3	147.6	4	2	1
338	PR 34641-2B-15-1-1-1	IRRI	FR 3/MALIGAYA SPECIAL # 6	385	625.2	833	135.7	1	2	1
339	IRGA 659-1-2-2-2	IRRI	IRGA 490/ECIA 31-18-11	396.8	608.4	792.9	142.4	5	2	1
340	CT 18160-3-2-2-3-1-M	CIAT	CT 11275-3-F4-8P-2/CAPIRONA//BCF1720	393.4	639.1	858.4	178.1	6	2	1

341	B11598C-TB-2-1-B-7	Indonesia	LIMBOTO//IRBL 23/IUF 5-1	397.4	568.8	687.2	103.2	NA	2	1
342	IR 04A216-B	IRRI	IR BB 60-1/IR 71730-51-2	532.7	800.6	1065	211.3	8	1	2
343	IRGA 318-11-9-2A	IRRI	NEWREX//IR 19743-25-2-2/BR IRGA 409	342.3	588.3	1025	210	10	2	1
344	IR 06A150-B	IRRI	IR02A127/IR 64	256	679.4	969.5	221.9	8	NA	NA
345	IR 06A144	IRRI	IR02A127/JANAKI	407	702.7	930.1	174.8	6	2	2
346	IR 77674-3B-8-2-2-8-2-AJY10	IRRI	IR 71730-51-2/IR 61920-3B-22-2-1 (NSIC RC 106)	494.2	690.4	863.6	141.9	4	2	1
347	IR 06A152-B	IRRI	IR01A130/JANAKI	490.2	651.6	837.6	111.6	2	2	2
348	PSBRC30	IRRI	IR 72/IR 24632-34-2	405.6	664.7	945	167.6	10	2	1
349	PR-116	PhilRice	PR 18/PAU 1628//PR 106	300.8	665.6	1127.5	277	10	2	1
350	CT 18148-6-9-3-3-2-MMP	CIAT	CT 9748-3-1-1P-2-M/CAPIRONA//BCF1720	434.7	715.8	1040	221.2	10	2	1
351	NSICRC122	IRRI	IR 44625-139-2-2-3/IR 32822-94-3-3-2-2	356.8	624.2	814.7	133.6	2	2	1
352	BALILLA	Italy	NA	418.2	691.8	1130	212.2	NA	NA	NA
353	WAS 122-IDSA-1-WAS-6-1 (NERICA-L-19)	AfricaRice	WAB 1291/3*IR 64	419.8	577.2	719.4	118.1	NA	2	1
354	IR 06N154	IRRI	IR 72158-11-5-2-3/IR 73707-45-3-2-3//IR 72875-94-3-3-2	333.2	743.7	928.6	202.3	8	2	1
355	BP1976B-2-3-7-TB-1-1	Indonesia	TB 47 K-MR-5/S 3385-5C-16-3-2//S 3385-5C-16-3-2///S 3385-5C-16-3-2	253.1	618.5	800.7	184.5	4	1	1
356	MINGHUI 63	China	IR30/Gui630	252.9	605.6	1045	267.4	10	2	1
357	IR 79195-42-1-3-1	IRRI	IR 72909-139-1-2-2/IR 68059-66-2-3-3-3	567	762.6	1120	185.1	10	2	1
358	CR 547-1-2-3	Egypt	IR 1615-31/IR 1605-64	231.1	529.7	880	201.4	NA	NA	NA
359	NSICRC158	IRRI	IR 73885-1-4-3-2-1-6 (MATATAG 9)/IR 70479-45-2-3//IR 64680-81-2-2-1-3	492.3	634.2	917.5	141.9	10	2	1
360	IR 65483-118-25-31-7-1	IRRI	IR 56*4/O BRACHYANTHA	286.8	583.2	967.5	194	10	2	1
361	CT 17323-1-1-2-2-2-2-M	CIAT	CT 16629-22/IR 25586-45-1-2//EPAGRI 108	518.4	646	795.3	105.8	8	2	1
362	PSBRC94	IRRI	IR 44535-22-3-3-3/IR 8866-30-3-1-4-2	363.2	567.5	952.5	199.3	10	2	1
363	B12743-MR-18-2-3	Bangladesh	PEPE/B 342 B-MR-1-3-KN-1-2-3-6-MR-3-BT-1	439.1	726.7	1030	234.9	5	2	1
364	OM 5629	Vietnam	C 27/IR 64//C 27	404	812.9	1085	215.1	8	1	1

365	IRI 346(PUNGSANBYEO)	IRRI	MILYANG 23/IR 1545	356.4	665	1087.5	244.1	10	2	1
366	IR 38	IRRI	IR 20*2/O NIVARA//CR 94-13	386.9	707.1	1282.5	279.5	10	2	1
367	HUA 565	China	NA	452.1	655.8	895	159.8	10	2	1
368	IR 05N229-B	IRRI	IR 74052-297-2-1/IR 72165-63-2-3-3	460.4	677.4	990	177.1	10	2	2
369	IR 05N419	IRRI	IR 72887-34-2-1-3/IR 73707-45-3-2-3	408.2	733.8	1078.4	224.5	8	1	2
370	CT 18657-2-1-2-1-2	CIAT	CT 18080-34(5)/ORYZICA 1	439.5	647.5	880	179.4	8	2	1
371	BRRI DHAN 28	Bangladesh	BR 6/PURBACHI	294.4	667.4	925	203.7	8	2	1
372	CT 17334-13-3-1-2-1-M	CIAT	PERLA/?//CT 10323-29-4-1-1-1T-2P///CT 10323-29-4-1-1T-2P/4/FEDEARROZ 50	373.7	554.5	875	156.9	10	2	1
373	PR 37126-PB-2-3-4-10-8	PhilRice	MALIGAYA SPECIAL # 6/IRBB 21	412.7	724.8	969.2	175.2	6	2	1
374	GUANG JIANG 1 (ACC 82336)	China	NA	322.8	565.8	1071.8	226.2	8	2	1
375	PEH-KUH-TSAO-TU (ACC 8237)	Taiwan	NA	372.4	648	805	155.8	NA	NA	NA
376	IR 03N137	IRRI	IR 73885-1-4-3-2-1-6 (MATATAG 9)/IR 70479-45-2-3//IR 64680-81-2-2-1-3	439.4	698.9	1042.5	185.7	10	1	2
377	IR 06N119	IRRI	IR 73707-45-3-2-3/IR 68552-100-1-2-2	351	620.8	745	139.4	6	1	1
378	PR 35789-B-37-3	IRRI	PR 30975-2-1/AR 32-19-3-3	375.4	624.8	871.8	145.5	6	1	1
379	IRGA 370-38-1-1F-C4-2	IRRI	ORYZICA 1/BR IRGA 412	216.8	509.7	733.6	203.5	4	2	1
380	PR 35766-B-24-3	PhilRice	PSB RC 14/PR 31561-AR 32-11-63-3	425.2	708.3	1152.5	213.5	10	1	1
381	CT 16658-5-2-2SR-2-3-6MP	CIAT	PERLA*2/ORYZA RUFIPOGON//CT 9748-13-2-1-M-M-1-1	473.2	749.2	1200	239	10	2	1
382	IR 79195-42-1-3-1	IRRI	IR 72909-139-1-2-2/IR 68059-66-2-3-3-3	383.4	793.2	1193.6	309.8	NA	1	1
383	IR 06N119-B	IRRI	NA	534.2	649.5	754.2	88.1	NA	2	1
384	IR 04N106-B	IRRI	NA	351.3	768.8	1255.7	335	NA	1	1
385	IR09M106	IRRI	DAESANBYEO/IR 65564-44-5-1	259	617.9	780	175.6	NA	1	1
386	PR37946-B-28-3-2	PhilRice	LP 368/PR 30646-3-5-17-MB-2-4-MB-1	421.7	716.2	922	172	NA	2	1
387	IR10N184	IRRI	IR 63896-60-3-1-2/IR 71676-106-10-3//IRRI 123	579.8	861.5	1235	206.3	NA	1	1
388	IR07N136	IRRI	IR 70114-5-3-3-3/IR 71718-59-1-2-3//IR 60912-93-3-2-3-	495	718.8	944.8	161.8	NA	1	1

			3							
389	PR37934-3B-8-2-1	PhilRice	C 7407 WH 11-3-3/PR 30595-B-2-1-1-1-2-4	586.7	996.3	1434.3	331.3	NA	1	1
390	PR37160-11-5-1-1-1-1	PhilRice	PJ 7//PSB RC 4/OKLAN	613.8	706.2	793.4	57.9	NA	1	1
391	IR10A107	IRRI	IR 66/IR 72870-19-2-2-3//BR 29	610.7	723.8	933.7	118.6	NA	2	1
392	IR05A235	IRRI	IR 73008-138-2-2-2/IR00A103	489	755.1	1102.6	192.2	NA	NA	NA

[a] Fixed names used in the International Rice Information System (IRIS).

[b] The breeding program or country.

[c] The end cross from which the genotype was selected.

[d] Min, minimum; Max, maximum; Sd, standard deviation.

[e] GG, Genotype group derived from cluster analysis based on environment standardized mean GY.

[f] NA, Not available.

[g] subpopulations based on SSR data analysis.

h subpopulations based on SNP data analysis

Table S2 Name, primer sequence and chromosome position of 53 random SSR markers.

Name	Forward primer	Backward primer	Chr.	Pos.(Mb)
RM495	AATCCAAGGTGCAGAGATGG	CAACGATGACGAACACAACC	1	0.2
RM581	ACATGCGTGATCAACAATCG	AATTGGATGTGGATGCACG	1	9.1
RM306	GTACGTAAACGCGGAAGGTGACG	CGACGTACGAGATGCCGATCC	1	12.3
RM113	CACCATTGCCCATCAGCACAAC	TCGCCCTCTGCTGCTTGATGGC	1	19.2
RM443	GCGAAGCCCAATCTGAAGAAGC	CCAGTCCCAGAATGTCGTTTCG	1	28.7
RM529	TTCACCACAACGATAGAGACTTCTGG	GGGAAGAAGATGACAGAGCAAGC	1	41.0
RM109	GCCGCCGGAGAGGGAGAGAGAG	CCCCGACGGGATCTCCATCGTC	2	0.2
RM322	CAAGCGAAAATCCCAGCAG	GATGAAACTGGCATTGCCTG	2	7.4
RM106	CGTCTTCATCATCGTCGCCCCG	GGCCCATCCCGTCGTGGATCTC	2	25.1
RM530	TTCTTTATTCCCTCGCACTGACC	CAATGATGCCACAAACCGTAACC	2	30.5
RM132	ATCTTGTTGTTTCGGCGGCGGC	CATGGCGAGAATGCCCACGTCC	3	1.0
RM554	GTTCGTCCGTCTCTCGTCTC	CCCAAAAATCTGTGCCTCTC	3	12.1
RM156	GCCGCACCCTCACTCCCTCCTC	TCTTGCCGGAGCGCTTGAGGTG	3	17.7
RM416	GGGAGTTAGGGTTTTGGAGC	TCCAGTTTCACACTGCTTCG	3	31.2
RM85	CCAAAGATGAAACCTGGATTG	GCACAAGGTGAGCAGTCC	3	36.3
RM551	CTTACTCCATTGGGCTGGAACC	TGTAGGGTGGTAAGAGATCCACTCC	4	0.2
RM261	CTACTTCTCCCCTTGTGTCG	TGTACCATCGCCAAATCTCC	4	6.6
RM437	ATCCCTCCTCTGCTCAATGTTGG	TCAGGGAGGGTCCTAGCTACTGG	4	8.3
RM185	GGCTCTCCATCTCCATTGATCC	GAGTTGTTGGGAGGGAGAAAGG	4	18.6
RM119	CATCCCCCTGCTGCTGCTGCTG	CGCCGGATGTGTGGGACTAGCG	4	21.2
RM127	CGAAGCTTTCGGTGGGATAGC	ACCTTGAGCGAGTCCTTGAACG	4	34.5
RM153	CCTCGAGCATCATCATCAGTAGG	TCCTCTTCTTGCTTGCTTCTTCC	5	0.2
RM330A	CAATGAAGTGGATCTCGGAG	CATCAATCAGCGAAGGTCC	5	6.6
RM173	CCTACCTCGCGATCCCCCCCTC	CCATGAGGAGGAGGCGGCGATC	5	21.6
RM188	TCCGCCTCTCCTCTCGCTTCCC	GCAACGCACAACCGAACCGAGC	5	22.7
RM136	GAGAGCTCAGCTGCTGCCTCTAGC	GAGGAGCGCCACGGTGTACGCC	6	8.8
RM3	ACACTGTAGCGGCCACTG	CCTCCACTGCTCCACATCTT	6	19.5
RM541	TATAACCGACCTCAGTGCCC	CCTTACTCCCATGCCATGAG	6	19.5
RM275	GCATTGATGTGCCAATCG	CATTGCAACATCTTCAACATCC	6	24.3
RM103	ATCAGCAGCATTCAGCATTTGG	CCGGACGATGTGTATATCTCTTGG	6	30.9
RM125	ATCAGCAGCCATGGCAGCGACC	AGGGGATCATGTGCCGAAGGCC	7	5.5
RM11	TCTCCTCTTCCCCCGATC	ATAGCGGGCGAGGCTTAG	7	19.3
RM336	GTATCTTACAGAGAAACGGCATCG	GGTTTGTTTCAGGTTCGTCTATCC	7	21.9
RM234	TTCAGCCAAGAACAGAACAGTGG	CTTCTCTTCATCCTCCTCCTTGG	7	25.5
RM172	TGCAGCTGCGCCACAGCCATAG	CAACCACGACACCGCCGTGTTG	7	29.6
RM152	AAGGAGAAGTTCTTCGCCCAGTGC	GCCCATTAGTGACTGCTCCTAGTCG	8	0.7
RM544	GCTGCACCCTCTCTCAATAAATGC	GTGGACAGCTCGAAACGAAGC	8	5.1
RM223	GAGTGAGCTTGGGCTGAAAC	GAAGGCAAGTCTTGGCACTG	8	20.5
RM256	GACAGGGAGTGATTGAAGGC	GTTGATTTCGCCAAGGGC	8	24.1
RM409	CCGTCTCTTGCTAGGGATTC	GGGGTGTTTTGCTTTCTCTG	9	14.4
RM434	TCTCTAGTTGCCTCATCCCTCTAACC	GGCTCAACCTCTATATTTGCTGATCG	9	15.7
RM245	ATGCCGCCAGTGAATAGC	CTGAGAATCCAATTATCTGGGG	9	22.3
RM222	CTTAAATGGGCCACATGCG	CAAAGCTTCCGGCCAAAAG	10	2.6

RM311	TGGTAGTATAGGTACTAAACAT	TCCTATACACATACAAACATAC	10	9.5
RM184	ATCCCATTCGCCAAAACCGGCC	TGACACTTGGAGAGCGGTGTGG	10	16.1
RM286	GGCTTCATCTTTGGCGAC	CCGGATTCACGAGATAAACTC	11	0.4
RM441	AAGGGAGTAGCCTCTCCATCTCC	GTGCTGACTCCTCTCCCTGTCC	11	6.1
RM287	GGCTACACCTACACGCGAGAACC	AGATGCATGGAATGCCTGTTTGG	11	16.7
RM457	GCACAAGTTGATACTCTCCTCTGACG	CCACCATTATCTGCTCCATCACC	11	19.0
RM144	CATGTTGTGCTTGTCCTACTGC	AGCTAGAGGAGATCAGATGGTAGTGC	11	28.3
RM247	AAGGCGAACTGTCCTAGTGAAGC	CAGGATGTTCTTGCCAAGTTGC	12	3.2
RM277	CGGTCAAATCATCACCTGAC	CAAGGCTTGCAAGGGAAG	12	18.3
RM17	TGCCCTGTTATTTTCTTCTCTC	GGTGATCCTTTCCCATTTCA	12	27.0